1943

Statistical investigation of a sample survey for obtaining farm facts

Raymond James Jessen
Iowa State College

NOTE TO USERS

STATISTICAL INVESTIGATION OF A SAMPLE SURVEY FOR OBTAINING FARM FACTS

by

Raymond James Jessen

A Thesis Submitted to the Graduate Faculty
for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: Agricultural Economics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State College

1943

TABLE OF CONTENTS

LIST OF TABLES

Table 28.2, other tables, and the questionnaires used are available in the printed copy of this dissertation. They were not reprinted in this manuscript copy since they did not bear directly on the problem of this thesis.

INTRODUCTION

This study was undertaken to add to our knowledge of the following questions pertinent to the problem of collecting data by the sample survey method.

(a) What is the amount and nature of error in interview data?

(b) What is the best available sampling procedure?

(c) By what method will sample data provide the best estimate of state or subdivision totals?

(d) How does cost affect the design of the sample?

The agricultural field survey is often used as a means of obtaining farm facts. In many cases it is regarded as superior to other methods of obtaining data, such as the mailed inquiry or the "cost route" (for example). But the field survey as it is ordinarily taken is quite unsatisfactory. Although most surveys are samples of some sort, this feature is frequently ignored by investigators. On this point the Social Science Research Council (1928) states that "little real significance is given to .. [sampling] at most [experiment] stations." This is not so much because investigators consider sampling as unimportant but that they fail to see that most of their investigations are samples if not by design then certainly by their use.[1] Before the sampling problems

---

[1] Investigators frequently will select a township or county for study because such a small area can be completely enumerated and thus the "unreliability" of sampling is avoided. It will usually be found, however, that these investigators, unwittingly perhaps, will either implicitly or explicitly generalize their findings to include some much larger population of which the small area examined is a small sample and often a very unrepresentative one at that.

of surveys can be adequately dealt with, it is first necessary to realize their sample character.

Another weakness of the field survey is that data obtained by interview are subject to errors resulting from imperfect memory, willful and unwillful mis-statements of the farmers and the carelessness or misunderstandings of the enumerators. These errors have long been realized but there appears to be a tendency among many investigators to either over- or underestimate their extent and magnitude and as a consequence the findings of studies based on survey data have frequently been wrongly condemned or praised. As a result, the effectiveness and efficiency of the survey method as a means for obtaining farm facts is not clear primarily because the method has not been adequately studied with these ends in view.

REVIEW OF LITERATURE

The following review of literature presents only what appear to be the more important studies of sampling problems which are relevant to the sampling of socio-economic data, particularly from farms. Studies appearing subsequent to the beginning of this investigation are included if they seem to fit into the general problem. The presentation is divided into the four kinds of problems previously mentioned.

(a) Amount and nature of error in interview data

Three kinds of enumeration errors (errors in interview data) may be distinguished: (a) questionnaire errors, that is, errors resulting from poor wording of the questions (either written or verbal); (b) respondent errors, that is, errors resulting from the failure of the respondent's memory to recall events accurately, from willful or unwillful mis-statements, misunderstanding, etc.; (c) enumerator errors, that is, errors resulting from carelessness or lack of skill on the part of the enumerator. Errors of the first kind have received considerable attention by persons sampling public opinions or attitudes. Blankenship (1940 A,B), Roper (1940) and Gallup (1940) have written on the problem of eliminating "loading" and ambiguity from questions. As an example the following suggestions for phrasing questions are made by Gallup (1940, p. 101).

1. The question should be as brief and to the point as possible. Long conditional or dependent clauses tend to confuse people.

2. The words or phrases should be simple and in common day-to-day use, among all groups in the community.

3. The questions should not include words which have a strong emotional content.

4. The questions must avoid all possible bias or suggestion in favor of or against a particular point of view.

5. The questions should include all the important alternatives which may emerge on a given issue.

6. Where the individual is being asked to choose between different alternatives, this choice of alternatives must be given as early in the question as possible.

7. In cases where the choices in a question are lengthy or numerous, it is preferable to list these on a card which the respondent can read. The average person is not likely to be successful in retaining a long list of alternatives, or complex questions, in his mind.

Little has been written in regard to errors attributable to the respondent. The "cash crop bias" mentioned by Sarle (1932) may be regarded as this sort of error. Morgan (1938) in a systematic study of enumeration errors points out their sources and suggests corrective measures. Recently Hopkins (1941) studying respondent errors of farm business items concluded they were both random and systematic and affect not only absolute values but also regressions.

Warren and Livermore (1911) suggested that enumerator errors could be reduced if enumerators would meet with their supervisor and transcribe the day's collection of data on listing sheets each evening, thereby detecting irregularities which can be corrected the next day either by phone or re-visit, while the memory is fresh. A large part

of the success of these pioneers of the survey method in this country appears to result from the thoroughness with which their enumerators were trained and supervised. The Social Science Research Council (1928) made some very pertinent remarks on the survey procedures followed in the late 1920's and offered some helpful suggestions for controlling the more common enumeration errors.

(b) Sampling procedure

Although sampling theory is probably as old as statistical theory only during the past 50 years has there been much progress in applying that theory in the collection of data. Kiaer (1895) of Norway and Bowley (1906,1913) of England were the leading early users and advocates of sampling for obtaining social and economic data. Renewed interest in sampling in the 1920's gave rise to a discussion of sampling methodology by Bowley (1926) and Jensen (1926 A,B) who were concerned mainly with the problem of how sampling units should be selected, that is whether purposively or by some scheme of random choice. This subject (what is called the two aspects of the representative method of sampling) was later studied by Neyman (1934) who concluded - with Bowley agreeing - that the stratified-random method of selection is superior to purposive selection except in what appear to be unusual instances. Gini's (1928) sampling of Italy's 1921 population census data is a well-known example of the purposive method.

During the 1930's interest in sampling methods in this country was stimulated by prospects of a sample census,[2] by increased demands for data on the part of the new action agencies of the government and by the introduction of improved methods of calculation. Some of the increased interest of this period can be seen by reading the proceedings of a conference sponsored by the BAE (1936) of the United States Department of Agriculture and Iowa State College held at Ames, Iowa, to discuss methods of sampling agricultural data. Many of the ideas suggested at this conference are still undergoing study.

Sampling progress since Neyman (1934) has been mainly along the lines of refinement although some new ideas have appeared. On stratification, Yates and Zacopanay (1935) suggest a convenient method by which the efficiency of stratification may be estimated by use of analysis of variance calculations. Recently Cochran (1939) presents further arithmetical suggestions for evaluating the efficiency of stratification, and King and Simpson (1940) and King and McCarty (1942) have given examples of the effectiveness of geographic stratification in sampling agricultural data.

The kind of sampling unit suitable for sampling socio-economic data is more of a problem when the smaller (but more efficient) randomly chosen units are used. The sampling unit adapted for this investigation, henceforth referred to as the grid method of sampling farms, is mentioned

---

[2]Sample censuses were being taken in Poland [Neyman (1934)], Bulgaria [Anderson (1934)] and Sweden [Hoijer (1937)].

in some of the papers of the BAE (1936) and appears in Holmes' (1939) study. In this case the sampling unit is defined as the farm or group of farms whose farmsteads are situated on the selected grid, the grid being some multiple or subdivision of the standard survey section. The sampling unit studied by King and Simpson (1940) for sampling acreages in crops appears to be similar but there is really an important difference. This unit which is defined as that area included within the boundaries of the selected grid is similar to the "quadrat" used by agronomists to sample experimental plot yields and by range management investigators to sample plant cover. With this scheme farms may or may not be enumerated as complete units according to whether or not their land area falls completely within the sample grid. Another type of sampling unit which has been used recently is what may be called a "random-point" scheme of selecting the units of observation. The standard survey section in this scheme provides a means for making a random selection of reference points (such as the intersections of the section's diagonals) in the vicinities of which a specified number of farms are designated for enumeration. This scheme was used by Finkner et al (1943) and probably was the principle followed by Mahalanobis (1940). The grid and aerial grid sampling units if properly used are unbiased, but the random-point procedure as it is usually followed is inherently biased. Finkner et al (1943), however, believe they have a method of removing this bias by their estimation procedure.

Statistical efficiencies of different sizes of sampling units suitable for agricultural sampling have not received much attention until recently. Neyman (1934) concluded that "it is much safer to make, say, 3,000 samplings of small units rather than 30 of larger ones, and this is true, whatever the stratification." Holmes (1939) reporting a study of farms in Morrow county, Ohio made by Stephan states that "it would require 15 per cent more farms in pairs, and 94 per cent more ... in strings of 8 [along a road] to make estimates as accurate as those secured by taking individual farms. With stratification these percentages were reduced to 11 and 64 per cent respectively." In his own study of the comparative statistical efficiencies of the (i) individual farm, (ii) single-section grid and (iii) four-section grid, he concluded that "the differences in accuracy between the ... [three sizes] are not highly significant." He does not have figures comparable to Stephan's. This absence of size effect, he states, "may be due to a high degree of success in stratification" and "the differences resulting from the two methods simply indicate that year-to-year change is not as highly correlated between neighboring farms as are the absolute figures for a single year." King and Simpson (1940) studied the effects of size and shape on the statistical efficiency of aerial grids in sampling crop acreages in 19 agricultural counties of the North Central Region and concluded that "in the aggregate the square unit was more efficient than the long narrow unit" and that

the single section grid was more efficient than either the four-section, sixteen-section or thirty-six-section grid.

Matching or pairing farms (or grids) may be a means of reducing the sampling error when change through time is being measured. It is a technique that has long been used by the BAE in its crop and livestock estimating work. The use of analysis of variance in evaluating the sampling efficiency of this method is concisely presented by Winsor and Clarke (1940).

Double sampling, that is an investigation in which two samples are taken, a small one to measure a hard-to-get character which is desired and a large one to measure some easy-to-get second character which is correlated with the first. Two kinds have appeared. In the method studied by Neyman (1938) the large sample is divided into groups designed to minimize the within-group variation of the easy-to-get character. The large sample subdivided in this manner can now be used as a stratified population from which to draw the small sample, or it can be used merely as a means for weighting the small sample. The other method is to use the small sample to determine the regression of the hard-to-get character on the easy-to-get character. The large sample may improve the estimated mean of the hard-to-get character by providing a more accurate estimate of the mean of the easy-to-get character in the regression equation. Watson (1937) used this method to improve an estimated mean leaf area of a large batch of leaves by double sampling for leaf weight, a much easier character to measure.

(c) Expansion or methods of estimation

Little is published which is relevant to the problem here. The BAE has done much work in this connection, most of which appeared as mimeographed material sent to their state statisticians. They have been concerned mainly with various ratios such as "ratio to land", "ratio to crop acres", "ratio relatives" which in most instances require a base such as the federal or state censuses. Yates and Zacopanay (1935) presented a procedure by which the efficiency of the ratio method of expansion can be estimated when the control is known without error. The case where the control is not known without error has been given by Cochran (1939) and by Schumacher and Chapman (1942). Procedures for obtaining the sampling variance of the case where the control is used for weighting subdivisions of the sample has been given by Cochran (1939) and Hendricks (1941). Recently Cochran (1942) described the conditions under which each of several kinds of ratio and regression methods of estimation are most efficient and gives comparisons of their relative efficiencies.

(d) Cost

The desirability of considering the cost of alternative sampling procedures has been recognized for some time by those setting up field plot designs. Zacopanay and Yates (1935) considered the problem of the most efficient balance between the percentage of an experiment sampled and the size of the experiment when the alternative amounts of work involved are taken into account. Watson's (1937) scheme of double

sampling to estimate leaf areas was devised to reduce the arduous labor involved in measuring areas of leaves when the simpler task of weighting leaves could be partially substituted for it. Neyman (1938) in a similar manner set up his double sampling scheme to determine the best use of sampling resources in sampling human populations. The problem with which Mahalanobis (1940) was concerned in sampling jute in India, was that of determining the best size of sampling unit for a given expenditure of survey funds. In this connection he presents some new ideas on the relationship of statistical efficiency and cost to size of sampling unit. Smith (1938) studied the relationship of statistical efficiency to size of sampling units used in field experiments and suggested a logarithmic function to describe it. In a study on sampling total wheat production, King and McCarty (1942) concluded that for a given expenditure it would be advisable to cover more of the wheat area by travel and sample fewer fields than they had done - a conclusion reached when the effects of cost were taken into account.

DESCRIPTION OF THE SURVEYS

To provide the data for the investigation, two sample surveys of Iowa farms were made during the last 2 weeks of December and the first 2 weeks of January, 1938-39 and 1939-40. The questionnaires used on these surveys carried questions designed to give general information on acreages and productions of crops; numbers of livestock bought, sold and on hand; receipts; expenditures and values of farm land and equipment; number of persons moving off and onto farms; number of cattle on feed, sow breeding plans, scale of living, etc. One of the objectives of the surveys was to test the feasibility of securing income information for a calendar year by a single interview, consequently many of the questions were constructed for this purpose.

Since the bulk of the information from the surveys applies to the calendar years 1938 and 1939, it will be convenient to use these year numbers to identify the two surveys. The sampling procedure of the 1938 survey was as follows: The quarter-section grid (an area of about 1/4 square mile or 160 acres) was selected for the sampling unit; the county was set up as the stratum. Townships were selected at random from each county - seven from an average size county of 16 townships, and in the same proportion for counties of other sizes.[3]

---

[3]Size of county was measured by the total number of agricultural quarter-sections it contains.

Quarter-sections were selected[4] at random from each of the selected townships - one each from five, two each from the remaining two.[5] An average size county would therefore have nine quarter-sections selected for the sample. The total number of agricultural quarter-sections in Iowa is about 219,176. Of these 908 (0.4 per cent of them) were selected for the sample.

The quarter-section grids merely designated the areas in which the farms to be enumerated would be found. A farm[6] was enumerated if its farmstead was situated within one of the selected grids. Farms were enumerated as complete units regardless of the location of their land area.

The 1939 survey was an integral part of the first survey. Four hundred and fifty-two (452) or approximately 50 per cent of the sample grids of the 1938 survey were selected for re-enumeration. Four hundred and forty-five (445) new grids were selected at random bringing the second sample up to a total of 897 grids. Both the old and the new grids were selected in the same manner as stated above. The reason for this particular sampling design will be discussed later.

---

[4]Only quarter-sections situated in an agricultural area were accepted. Those situated in incorporated town limits, lakes, rivers or national parks were rejected and new selections made.

[5]The selections were made in this manner in order to assure that a sufficient number of townships would contain two sample quarter-sections with farms. This would permit of better estimates of quarter-section variance within townships. Later, however, it was found that randomizing quarter-sections within counties (ignoring townships) would have given about the same geographical distribution and therefore the above precaution was unnecessary.

[6]A tract or tracts of land 3 acres or more under one management was considered a farm. This followed the Iowa assessors definition. See footnote 8 for elaboration on this point.

Enumeration Procedure

Enumerators were instructed to visit each farmstead situated on the selected grids to interview either the operator or whoever might be familiar with the farm's business. If no one was found at home or if non-cooperation was met, the enumerator was instructed to substitute the next nearest farm in that vicinity, a record being kept of all such cases.

Instructions for grids selected for revisitation were as follows:

(a) Change of operator and change in farm acreage were ignored.

(b) New farms (those appearing since the previous visit) were to be enumerated.

(c) Farms disappearing were recorded.

(d) If during the first visit farms were substituted these were re-enumerated instead of those for which substitution was made.

Enumerators were in the field during the last 2 weeks of December and the first 2 weeks of January. The bulk of the enumeration, however, was accomplished during the middle of this period.

General Discussion on the Questionnaires and on Field Operations

The time required for enumerating the questionnaires was 32 minutes for the 1938 questionnaire and approximately 50 minutes for that of 1939. The enumerators felt that the longer questionnaire was nearing the maximum desirable for this kind of an inquiry where little was done to

acquaint the farmers beforehand of what was to be asked of them. In an opinion poll of the 15 enumerators employed on the 1939 survey, it was found that they unanimously agreed that letters sent to farms selected for revisitation (matched farms) were helpful. (These letters were sent out about a week before enumeration began and contained a brief statement thanking the farmer for his cooperation last year and a statement that we intended to revisit him this year.) Most enumerators were of the opinion farmers would appreciate something in exchange for their effort - experiment station or government publications, a report on the findings of the inquiry, etc.

Finding the designated quarter-sections in the field was not very difficult except in the northeastern section of the state and in other isolated instances. In most of the northeastern counties it was found advisable first to visit the county AAA offices where their aerial maps were used to secure information such as number of farms, if any, landmarks, etc., useful to the enumerator.

THE DATA

Usable records were obtained from 773 farms in 1938 and 782 farms in 1939. In the following tables are presented state estimates for selected items based on data from the two sample surveys compared with figures supplied by the State Farm Census (taken by the township assessor), AMS, and Federal Census.

It can be seen that for some items the sample surveys compare quite favorably with other sources of data, whereas for others there appear to be serious differences. An investigation of this matter will be presented below under the topic "Errors".

Errors Due to the Vagaries of Sampling

Estimates of the standard errors of the sample means, expressed as per cent of the mean, have been computed for a selected group of items, on both an individual farm and quarter-section grid basis. These appear in Table 5.

It can be seen that the degree of precision by which various farm items can be sampled varies rather widely. For the items in the table, relative sampling errors range from 0.7 to 17.1 per cent for the farm mean and from 2.4 to 14.2 per cent for the grid mean. Most of the more important items, however, are around 3 per cent or 4 per cent and are slightly less on an individual farm basis.

Table 1. Estimates of Totals for the State of Iowa on 28 Farm Items Obtained from Three Sources: (a) Sample Survey, (b) Iowa State Farm Census and (c) AMS 1938 and 1939

	1938-Source of data			1939-Source of data		
	Sample survey[a] (thousands)	State census[b] (thousands)	AMS[c] (thousands)	Sample survey[a] (thousands)	State census[d] (thousands)	AMS[e] (thousands)
Land in farms (acres)	34,080	34,403	-	34,080	34,545	-
Acreages (harvested)						
Total corn	10,149	10,270	10,417	9,272	9,373	9,688
Husked	9,557	9,709	9,844	8,832	8,943	9,261
Silage	210	233	240	170	179	194
Fodder	198	211	-	160	156	-
Hogged	184	116	-	110	95	-
Oats, grain	5,980	5,923	5,972	4,838	4,973	5,076
Wheat, grain, winter and spring	635	581	592	426	389	390
Barley, grain	393	422	447	587	544	563
Soybeans, grain	331	306	321	572	539	487
Alfalfa, hay	898	814	879	845	789	879
Production of crops						
Corn, grain (bu.)	455,550	449,509	452,824	481,353	467,055	481,572
Oats, grain (bu.)	206,753	206,205	209,020	149,954	154,159	154,818
Wheat, grain, winter and spring	9,152	9,091	9,284	6,432	6,726	6,490
Barley, grain (bu.)	12,266	12,831	13,634	13,540	12,533	13,794
Soybeans, grain (bu.)	6,856	6,462	6,741	11,738	11,096	10,227
Alfalfa, hay (tons)	1,895	1,797	1,934	1,636	1,657	1,846
Livestock (1/1/39: 1/1/40)						
Horses, all ages (head)	763	723	783	743	-[h]	752
Mules, all ages (head)	55.8	47	55	45.8	-[h]	54
Cattle, all ages (head)	4,295	4,001	4,465	4,721	-[h]	4,688
Sheep, all ages (head)	1,303	1,229	1,710	1,105	-[h]	1,844
Swine, all ages (head)	7,398	6,512	8,179	10,240	-[h]	9,651
Chickens (head)	28,661	27,377	30,172	31,736	27,846	30,930
Miscellaneous						
Sows bred or to be bred for spring farrow (head)	1,765	1,707	1,643[f]		1,608	1,778[g]
Number of tractors	122	111	-		118	-
Number of autos	189	190	-		-	-
Number of trucks	20.2	19.8	-	25.1	18.8	-
Number of radios	158	151	-	178	-	-

[a] Estimated from sample survey data by expansion method 2 for the rural (unincorporated) area of Iowa only.
[b] From the Thirty-Ninth Annual Iowa Yearbook of Agriculture. Figures rounded.
[c] Acreages and production of crops taken from Crop Report release of Dec. 19, 1939. Livestock figures taken from Livestock Report released Feb. 15, 1940.
[d] From the Fortieth Annual Iowa Yearbook of Agriculture. Figures rounded.
[e] Same source as (c). Preliminary.
[f] Crops and Markets, December, 1938.
[g] Livestock Report of Dec. 22, 1939.
[h] Collection of data on these items was discontinued in 1939.

Table 2. Yields of Harvested Crops Obtained from Sample Survey, State Census and AMS, State of Iowa, 1938 and 1939

Crop	1938			1939		
	Sample Survey[a]	AMS preliminary[b]	State census[c]	Sample survey[d]	AMS preliminary[e]	State census[f]
Corn, bu./acre	47.7	45.5	46.3	54.5	52.0	52.2
Oats, bu./ acre	34.6	33.5	34.8	31.0	30.5	31.0
Wheat, all, bu./acre	14.4	16.4	15.7	15.1	16.6	17.3
Barley, bu./acre	31.3	29.0	30.4	23.1	24.5	23.0
Soybeans, bu./acre	20.7	19.5	21.1	11.5	14.5	14.1
Alfalfa, hay, tons/acre	2.11	2.20	2.21	1.94	2.10	-
Soybean hay, tons/acre	-	-	-	2.06	1.50	-
Clover and timothy hay, tons/acre	-	-	-	1.28	1.05	-

[a] Data from the 1938 survey of 773 farms.
[b] Crops and Markets, Vol. 15, No. 12, December, 1938.
[c] Thirty-Ninth Annual Iowa Yearbook of Agriculture (1939).
[d] Data from the 1939 survey of 782 farms.
[e] General Crop Report, Dec. 19, 1939.
[f] Fortieth Annual Iowa Yearbook of Agriculture (1940).

Table 3. Estimates of Selected Farm Receipt Items[a] for the State of Iowa, Sample Surveys and AMS, 1938 and 1939

Item	1938		1939	
	Sample survey	AMS[b]	Sample survey	AMS[b]
	($1,000)	($1,000)	($1,000)	($1,000)
1. Government payments	22,769	29,719	55,214	69,444
2. Value of home-used livestock and livestock products		33,550	24,040	28,365
3. Receipts from sales of:				
a. Cattle, calves, beef and veal	106,088	145,316	135,484	161,402
b. Hogs, pork and lard	114,553	190,393	140,702	167,994
c. Sheep, lambs, mutton and lamb	3,853	9,855	5,924	11,051
d. Chickens		19,275	7,935	17,486
e. Eggs	16,509	27,653	18,116	23,868
f. Dairy products	48,495[c]	65,928	51,735[c]	60,789

[a]Items are not strictly comparable. AMS figures do not include inter-farm transactions, whereas the sample survey figures do. The actual discrepancies therefore are somewhat larger than they appear in this table. Sample survey estimates are based on the raw data - no adjustments for bias have been made.

[b]Preliminary.

[c]A net figure. Products brought back have been deducted.

Table 4. Percentage of Iowa Farms in Each Tenure Group; Data from the 1939 Sample Survey and the Federal Census

Tenure group	Sample survey 1939	Federal census	
		1935	1940
	(%)	(%)	(%)
Owner	37.3	39.2	41.3
Renter	50.9	49.6	47.6
Part-owner	11.3	10.5	10.5
Manager	0.5	0.7	0.6

Table 5. Estimated Relative Sampling Errors of Selected Items on Both a Per Farm and Per Quarter-Section Basis, 1938 and 1939 Surveys

Item	Individual farm basis		Quarter-section basis
	1938	1939	1939
	(%)	(%)	(%)
1. Number of swine	3.8	3.0	4.1
2. Number of horses	3.1	3.4	3.5
3. Number of sheep	14.4	9.6	15.0
4. Number of chickens	2.2	2.5	3.5
5. Number of eggs yesterday	3.8	4.3	5.0
6. Number of cattle	3.6	2.7	4.1
7. Number of cows milked yesterday	2.5	2.7	3.6
8. Number of gallons milked yesterday	3.1	3.2	
9. Receipts from dairy products	4.1		5.0
10. Farm acres	1.9	2.0	3.3
11. Corn acres	2.6	2.2	3.5
12. Oat acres	3.2	2.7	3.7
13. Corn yield	1.1	0.7	
14. Oat yield	1.3	1.6	
15. Feed expenditures, farm	8.6		6.0
16. Total expenditures, operator	5.2	3.3	
17. Total receipts, operator	5.6	3.3	4.1
18. Net cash income, operator	5.9	9.0	8.7
19. Corn sealed, operator	13.1		
20. Government payments, farm	4.9		
21. Number of hogs sold, farm		3.4	4.2
22. Number of cattle sold, farm		9.1	6.9
23. Number of hogs bought, farm		17.1	13.9
24. Number of cattle bought, farm		11.8	10.2
25. Number of cows milked during the year		2.3	
26. Number of hens and pullets, laying age		2.5	
27. Net income, operator			4.1
28. Number of persons on farm			3.2
29. Number of autos			3.0
30. Number of farms			2.8

In Table 5a are shown estimated sampling errors of the 1938 survey on a type-of-farming area basis together with those for the state as a whole. Since these sampling errors have been estimated from sample data they are themselves subject to the vagaries of sampling. In view of this, therefore, differences in estimated sampling errors among type-of-farming areas may or may not be real differences and must be interpreted with some caution. Some interesting conclusions, however, can be drawn from these data. In the following items: numbers of swine, horses, cattle; farm acres, corn acres, corn yield and feed expenditures, the Southern Pasture Area had the highest sampling errors. For number of sheep it had the lowest. This is useful information if type-of-farming area inquiries are to be made. Under such circumstances samples for general inquiries in the Southern Pasture Area should be somewhat larger than those for other areas since many of the important items show higher variability there.

The difference between the two sets of sampling errors is due in part to the variations of sampling and in part to a real difference in the variabilities of items taken on the two bases.

If $\bar{x}_g$ is the sample mean of an item on a grid basis

$\bar{x}_f$ is the sample mean of an item on a farm basis

$\bar{f}_g$ is the sample mean of number of farms on a grid basis

then
$$\bar{x}_g = \bar{x}_f \bar{f}_g \qquad (1)$$

Table 5a. Estimated Relative Sampling Errors of Selected Items from the 1938 Survey by Type-of-farming Area and for the State

Item	Standard error as a per cent of the mean					
	Northeast dairy area	Cash grain area	Western livestock area	Southern pasture area	Eastern livestock area	State of Iowa
	(%)	(%)	(%)	(%)	(%)	(%)
(Number of farms)	(141)	(158)	(169)	(143)	(162)	(773)
Number of swine	7.2	7.7	7.8	11.3	9.0	3.8
Number of horses	5.5	5.8	5.6	6.3	5.7	3.1
Number of sheep	31.8	40.1	30.0	17.6	24.1	14.4
Number of chickens	5.8	4.7	4.6	5.8	4.3	2.2
Number of eggs yesterday	9.0	8.5	8.8	8.9	7.5	3.8
Number of cattle	6.6	6.7	7.8	10.6	8.9	3.6
Number of cows milked yesterday	5.5	5.2	7.6	5.6	5.5	2.5
Number of gallons milked yesterday	6.7	6.6	6.1	8.1	7.2	3.1
Receipts from dairy products	6.7	10.8	9.9	9.8	7.5	4.1
Farm acres	4.9	4.1	4.4	7.1	5.1	1.9
Corn acres	5.4	4.7	5.2	7.8	5.4	2.6
Oat acres	5.9	4.8	9.1	9.2	5.9	3.2
Corn yield	2.3	1.7	2.5	2.9	2.5	1.1
Oat yield	2.8	2.2	2.5	2.2	3.6	1.3
Feed expenditures, farm	13.0	11.2	12.1	26.3	22.3	8.6
Total expenditures, operator	11.1	10.0	9.8	10.7	13.7	5.2
Total receipts, operator	8.3	[a]	9.3	10.4	11.2	5.6[b]
Net cash income, operator	8.8	[a]	15.1	12.5	11.3	5.9[b]
Corn sealed, operator	56.1	19.7	23.9	40.4	28.8	13.1
Government payments, operator	13.6	8.7	11.0	10.4	8.9	4.9

[a] Not available.
[b] Cash Grain Area not included in estimate of variance.

Now the variances must be the same in both cases, therefore,

$$\sigma^2_{\bar{x}_g} = \sigma^2_{(\bar{x}_f \bar{f}_g)} \qquad (2)$$

$$\sigma^2_{\bar{x}} \doteq \left(\bar{x}_f \bar{f}_g\right)^2 \left(\frac{\sigma^2_{\bar{x}_f}}{\bar{x}_f^2} + \frac{\sigma^2_{\bar{f}_g}}{\bar{f}_g^2} + 2\rho \frac{\sigma_{\bar{x}_f}}{\bar{x}_f} \frac{\sigma_{\bar{f}_g}}{\bar{f}_g} \right) \qquad (3)$$

or $$V^2_{\bar{x}_g} \doteq \left(V^2_{\bar{x}_f} + V^2_{\bar{f}_g} + 2\rho V_{\bar{x}_f} V_{\bar{f}_g} \right) \qquad (4)$$

where V is the relative standard error of the sample mean and ρ is true correlation of $\bar{x}_f$ and $\bar{f}_g$.

It can be seen in (4) that if ρ (that is, the correlation of item mean per farm with mean farms per grid) is zero then

$$V^2_{\bar{x}_g} \doteq V^2_{\bar{x}_f} + V^2_{\bar{f}_g} \qquad (5)$$

Hence differences in the relative sampling errors of the two sets of means are due in part to the relative variance of farms per grid and the correlation of item mean per farm with mean farms per grid. It cannot be said with certainty, therefore, that the differences shown in the table reflect real differences in sampling efficiency. More on sampling efficiency will be considered later. The point here is merely to note the order of magnitude of sampling errors on both farm mean and grid mean basis and to show roughly how they might be related.

Methods of Estimating State Totals from the Sample

Only three of the available methods of expansion were investigated in this study. Knowledge of the total number of quarter-section grids, total land in farms and total number of farms in the state provide the basic data for making expansions. In our case the quarter-section count can be made from a suitable set of maps (see appendix D, pp. 95 to 97) and figures on the total land in farms and total number of farms are available in the Iowa State Farm Census reports and the Federal Census. With these quantities known it is a simple operation to derive estimates of totals from a sample. For example, it has been found that there are 224,180 quarter-section grids in rural Iowa. Using this as a multiplier, item means per grid in the sample can be expanded to state totals. Similarly, expansions can be based on total farm acres or total number of farms. The three methods are presented symbolically in the following paragraphs.

If the total number of quarter-sections in the state is denoted by Q, the item mean per quarter-section (given by the sample) by $\bar{x}$ and the estimated state total of the item by $\hat{X}$, then

(Method 1) $$\hat{X} = Q\bar{x} \qquad (6)$$

If A is the total acres of land in farms, F the total number of farms in the state, $\bar{a}$ the mean acres in farms per quarter-section and $\bar{f}$ the mean number of farms per quarter-section, then also

(Method 2) $$\hat{X} = A\frac{\bar{x}}{\bar{a}} \qquad (7)$$

and

(Method 3) $$\hat{X} = F\frac{\bar{x}}{\bar{f}} \qquad (8)$$

To show the relationships that (methods 2 and 3) have to (method 1), these formulas can be written:

$$\hat{X} = Q\bar{x}\left(\frac{A/Q}{\bar{a}}\right) \qquad (7a)$$

and

$$\hat{X} = Q\bar{x}\left(\frac{F/Q}{\bar{f}}\right) \qquad (8a)$$

wherein (methods 2 and 3) become merely methods for adjusting (method 1), according to whether or not the sample deviates from the true values of the two characters, farm acres or number of farms per quarter-section. (Methods 2 and 3) require that A/Q and F/Q be known from sources other than the sample, such as a state or federal census for example. Before we can properly determine which of these methods provides the best estimate of state totals we should first consider the conditions under which each is appropriate.

The discrepancies in the figures for farm acres per quarter-section may be due to any one or more of the following three causes:

(a) Bias; for instance, enumerators have failed to account for every farm situated on the selected quarter-sections.

(b) Quarter-sections have been selected which have fewer than average number of farms - a chance occurrence of the random sampling of quarter-sections.

Table 6. Farm Acres and Number of Farms per Quarter-Section as Indicated by State Census and Sample Survey (1938 and 1939) Data: and their Standard Errors

Item	State census (adjusted[a])	Sample survey	
		1938	1939
Farm acres per quarter-section	155.5	154.3	151.2
Standard error	(0.23)[b]	[c]	(4.8)
Number of farms per quarter-section	0.9159	0.8721	0.8628
Standard error	(0.010)[b]	[c]	(0.022)

[a] 1938 report. Figures adjusted to remove incorporated areas. See appendix E, pp. 97 to 104.
[b] Estimated from the first differences of the time series including the period 1925-1937.
[c] Standard errors were not computed for 1938 data. Should be approximately the same as for 1939.

(c) A variant of (b) where, although the quarters chosen were those having the average number of farms situated on them, these farms were less than average in size. This also could be the result of sampling variation.

Causes (b) and (c) are probably independent of (a) but are positively correlated with each other (correlation of total farm acres by number of farms on quarter-sections was +0.71). Both sample surveys taken individually appear to agree quite well with the state census figure for farm acres per quarter in view of the sampling error. Taken together, however, the two surveys show signs of a downward bias. The farms per quarter figures show the same tendency. This was to be

expected because of certain decisions governing enumeration procedure.[7]

We have little evidence on the effects causes (b) and (c) may have had on the discrepancies under consideration. The following data may help to show the information we do have.

Table 7. Farm Acres per Farm as Given by State Census (1938) and Sample Survey (1938 and 1939) Data, and Standard Errors

Item	State census (adjusted)	Sample survey	
		1938	1939
Farm acres per farm	169.7	176.9	175.2
Standard error	-	3.4	3.5

The sample survey farms appear to be larger than those reporting to the state census. Does this mean that the reverse of cause (c) has taken place - that quarters having farms larger than average farms were selected? Not necessarily. First, because for the sample survey a

[7](1) Only those quarter-sections were visited where we had some evidence that at least one farmstead was situated on them, the evidence being the information available on soils maps which were not accurate for the present situation, hence farms which might have been existing on the unvisited quarter-sections were never given the opportunity of being counted; (2) if errors are made in counting the farms on the visited quarter-sections it seems reasonable to expect that they are more likely to be the result of farms being overlooked rather than that of farms being counted which really were not situated within the confines of the selected quarter-section.

farm was defined so as to approximate an operating unit,[8] and therefore would tend to be larger than that of the state census, and secondly, because there may be bias arising from the method of substituting farms where information on the originally selected farms were not available. However, there is no evidence of substitution bias[9] and there seems to be reason enough to believe that the sample survey farm size is really larger than that of the state census. Consequently, the sample survey figures in Table 6 are not directly comparable with those of the state census. Our best estimate of average size of farm where a farm is defined as in the sample survey is that indicated by sample survey data. Hence, with no available check-data, we are unable to determine whether we are above or below the true value and therefore cause (c) must be rejected because of lack of evidence. Similarly, cause (b) must be rejected.

---

[8] Although both the state census and sample survey did not include tracts of land less than 3 acres as farms, it is not clear how state census enumerators consider tracts having complex control. For instance, during sample survey field operations cases were found where perhaps a father exercising complete control over 160 acres, had a son farming an 80-acre tract as his own but who used his father's machinery and equipment, lived with his father, and perhaps served as a hired hand on his father's farm. In such instances, if no clear-cut transactions were carried on between father and son the two tracts were considered as belonging to a single operating unit under the joint control of father and son. Frequently the control of brothers was found to be most easily handled by combining all operations into that of one "farm". Cases where tracts would be operated as farms, although no buildings were located on them, were not found although several farms were found the operators of which lived in town. Separate tracts, even if widely separated, were considered as parts of a single farm if it appeared that they were operated as part of a larger enterprise. This is merely evidence given to show why the farm defined as an operating unit would probably be larger than the farm as defined by the state census.

[9] See pp. 46 to 49.

Let us again consider the discrepancies in farm acres per quarter as shown in Table 6. We concluded that at least a part of this could be explained as the result of a bias. A comparison of the standard errors of these figures indicates that the state census figure is by far the more precise (it must be remembered that the use of a standard error on the state census figure is a crude one but probably useful in the sense in which we shall use it). Total land in farms in Iowa varies relatively little from year to year. It seems reasonable, then, to accept the state census figure of 155.5 as the best estimate of farm acres per quarter-section (for both years - the change is negligible). The most reasonable method of adjusting the sample survey data, then, in view of the probable bias, is to assume the total discrepancy (for both years) to be that of bias, and therefore to multiply (method 1) estimates by the factors 155.5/154.3 or 1.008 and 155.5/151.2 or 1.029 for the years 1938 and 1939, respectively (method 2). Since accurate figures on number of farms (as defined by the sample survey) are not available, we shall not attempt to estimate by (method 3). We shall later consider its potential precision, however.

Comparative Precision of the Three Methods of Estimation

In the foregoing discussion we were concerned with the discrepancies between the two sample surveys and the state census on the quantity, number of farm acres per quarter-section, and concluded that this could well be due to a bias and with this being the case, that (method 2) was a proper method by which state estimates could be made.

We are concerned now with the problem of determining the variances associated with each of these methods. Approximations are given by the following formulas:

$$\sigma^2_{\hat{\bar{X}}_{(1)}} = Q^2\bar{x}^2\left(v^2_{\bar{x}}\right) \qquad (9)$$

$$\sigma^2_{\hat{\bar{X}}_{(2)}} = Q^2\bar{x}^2\left(v^2_A + v^2_{\bar{x}} + v^2_{\bar{a}} - 2v_{\bar{x}}\, v_{\bar{a}}\, r_{\bar{x}\bar{a}}\right) \qquad (10)$$

$$\sigma^2_{\hat{\bar{X}}_{(3)}} = Q^2\bar{x}^2\left(v^2_F + v^2_{\bar{x}} + v^2_{\bar{f}} - 2v_{\bar{x}}\, v_{\bar{f}}\, r_{\bar{x}\bar{f}}\right) \qquad (11)$$

where v = coefficient of variability $= \frac{\sigma}{m} = \frac{\text{standard deviation.}}{\text{mean}}$

It is apparent that when

$$r_{\bar{x}\bar{a}} > \frac{v^2_A + v^2_{\bar{a}}}{2v_{\bar{x}}\, v_{\bar{a}}} \qquad (12)$$

and likewise when

$$r_{\bar{x}\bar{f}} > \frac{v^2_F + v^2_{\bar{f}}}{2v_{\bar{x}}\, v_{\bar{f}}} \qquad (13)$$

the variances of (methods 2 and 3) will be smaller than that of (method 1).

As a first approximation, let us assume that the total number of farms and total land in farms (F and A) are known without error (that is, v^2_F and $v^2_A = 0$).

The variances of (methods 2 and 3) relative to (method 1) will be

$$\frac{\sigma^2_{\bar{X}\,(2)}}{\sigma^2_{\bar{X}\,(1)}} = 1 + \frac{v^2_{\bar{a}}}{v^2_{\bar{x}}} - 2\frac{v_{\bar{a}}}{v_{\bar{x}}} r_{\bar{x}\bar{a}} \qquad (14)$$

and

$$\frac{\sigma^2_{\bar{X}\,(3)}}{\sigma^2_{\bar{X}\,(1)}} = 1 + \frac{v^2_{\bar{f}}}{v^2_{\bar{x}}} - 2\frac{v_{\bar{f}}}{v_{\bar{x}}} r_{\bar{x}\bar{f}} \qquad (15)$$

The reciprocals of these will be a measure of relative precision. Taking (method 1) as a standard (precision = 100) the relative precisions of (methods 2 and 3) were computed for a number of items and appear in Table 8.

It is clear that if total acres in farm land is known, (method 2) is in general, the most precise method of expanding sample data. For the items: number of persons on farms, number of automobiles and number of chickens, however, (method 3) is best. Unfortunately, the total number of farms in a state at a given time is generally not known accurately. If we accept rough estimates based on time series data as measurements of the precision of these quantities, we find that v^2_A = 0.00000225 and v^2_F = 0.00011264. Including those elements of variation in the variances of the three methods we have the comparisons which appear in Table 9.

We conclude from Table 9 that variation in the total land in farms from year to year in Iowa does not greatly affect the precision of

Table 8. Precision of Expansion (Methods 2 and 3) Compared with (Method 1) for Selected Items of the 1939 Survey. Total Land in Farms and Total Number of Farms Assumed to be Known Without Error

Item	(Method 2) as a per cent of (Method 1)	(Method 3) as a per cent of (Method 1)
Corn acres, harvested for grain	382	170
Number of cattle	234	140
Oat acres, harvested for grain	220	-
Number of swine	205	148
Number of hogs sold, 1939	192	-
Number of persons on farms	181	261
Number of horses	169	-
Total receipts, operator	156	-
Net income,[a] operator	148	106
Number of automobiles	147	702
Number of cows milked yesterday	137	-
Commercial feed expenditures, farm	129	128
Number of chickens	123	203
Number of cattle sold, 1939	120	-
Number of cattle bought, 1939	113	-
Number of farms	113	-
Receipts from dairy products, farm	112	125
Number of eggs yesterday	112	-
Number of sheep	106	-
Net cash income, operator	103	106
Number of hogs bought, 1939	102	-

[a]Includes an allowance for changes in inventory.

Table 9. Precision of Expansion (Methods 2 and 3) Compared with (Method 1) for Selected Items of the 1939 Survey. Total Land in Farms and Total Number of Farms Assumed Estimated from Time Series Data

Item	(Method 2) as a per cent of (Method 1)	(Method 3) as a per cent of (Method 1)
Corn acres harvested	379	146
Number of cattle	234	128
Oat acres, harvested for grain	219	-
Number of swine	205	118
Number of hogs sold, 1939	192	-
Number of persons on farms	180	208
Number of horses	169	-
Total receipts, operator	156	-
Net income,[a] operator	148	99
Number of automobiles	147	367
Number of cows milked yesterday	137	-
Commercial feed expenditures, farm	129	122
Number of chickens	123	168
Number of cattle sold, 1939	120	-
Number of cattle bought, 1939	113	-
Number of farms	113	-
Receipts from dairy products, farm	112	118
Number of eggs yesterday	112	-
Number of sheep	106	-
Net cash income, operator	103	105
Number of hogs bought, 1939	102	-

[a]Includes an allowance for changes in inventory.

(method 2). Variation in total number of farms as found by the Iowa state census does have a notable effect on the precision of (method 3). Even after allowance has been made for error in estimating the controls, total land in farms and total number of farms, both (methods 2 and 3) are usually more precise than (method 1), and in some cases these gains are rather substantial.

It is interesting to note that no great improvement can be made in estimating total number of farms by knowing total land in farms (the increase in efficiency is 13 per cent). This suggests that increasing the number of quarters in the sample by 13 per cent would give by (method 1) the same precision as the smaller sample using (method 2).

Measuring Year-to-year Differences and Percentage Changes

We wish to compare the relative sampling efficiencies of two methods of measuring year-to-year differences; that is, by samples drawn independently each year and by a matched sample. Data for the matched sample were provided by those quarter-sections which were visited both years. The problem with which we are here concerned is the estimation of the sampling errors of the year differences which each of these sampling procedures propose to measure. The computations can be most conveniently carried through in the form of an analysis of variance, given in Table 10.

Table 10. Analysis of Variance of Swine Inventories on 452 Matched Grids

Source	Degrees of freedom	Mean square	Mean square an estimate of
Total	903		
Years	1	31,308	
Counties	99	4,979	
Grids within counties	352	3,913	$\sigma^2_{YG} + 2\sigma^2_G$
Year x county	99	627	
Year x grids within counties	352	396	σ^2_{YG}

Total number of swine, Jan. 1, 1939 14,583
Total number of swine, Jan. 1, 1940 19,903

Difference (increase) = 5,320 or 11.8 swine per grid.

Following the suggestions of Winsor and Clarke (1940) we shall assume that the number of swine situated on a given grid for a given year is composed of the components:

(a) A mean for all grids for all years.

(b) A deviation due to year, common to all grids.

(c) A deviation due to county, common to all grids within the county but varying from county to county.

(d) A deviation due to the grid, common to all years but varying from grid to grid.

(e) A residual deviation, affecting each grid independently.

Let us denote the variance of components d and e, respectively, by σ^2_G and σ^2_{YG}. Furthermore let us assume that the deviations d and e are independent and random. We wish to draw samples stratified by county which will provide estimates of the population difference between two given years.

Two methods of sampling are to be compared. First, an unmatched sample, that is, one in which grids are selected at random within each county independently in each of the 2 years. The error variance of the year difference, per grid, in this case will be given by

$$2\left(\sigma^2_G + \sigma^2_{YG}\right)$$

Second, a matched sample, that is, one in which a single set of grids is selected at random within each of the counties and is used for both years. The error variance of the year difference, per grid, in this case will be given by

$$2\sigma^2_{YG}$$

The mean squares in the analysis of variance table given above are estimates of these quantities where:

Mean square of grids within counties is an estimate of

$$2\sigma^2_G + \sigma^2_{YG}$$

Mean square of Y X G within counties is an estimate of

$$\sigma^2_{YG}$$

The comparative efficiencies of the matched versus the unmatched samples for measuring year differences (stratified by county) will be given by the ratio

$$\frac{2\left(\sigma_G^2 + \sigma_{YG}^2\right)}{2\sigma_{YG}^2} \text{ or,}$$

$$\frac{\text{mean square of grids within counties} + \text{mean square of Y} \times \text{G within counties}}{2(\text{mean square of Y} \times \text{G within counties})}$$

which gives the number of pairs of unmatched grids which are equivalent (give same sampling precision) to one matched grid.

For swine,

$$\frac{2\left(\sigma_G^2 + \sigma_{YG}^2\right)}{2\sigma_{YG}^2} = \frac{3{,}913 + 396}{2(396)} = 5.4$$

In Table 11 comparative efficiencies are shown for a number of items.

It is quite evident that substantial gains are obtained by matching, although much variation exists among items.

Discussion - a digression

The analysis of variance set forth in Table 10 contains in addition to that which was just discussed, much interesting information. For instance, a simple test of statistical significance is provided by the mean squares for years and Y × G within counties. In this case $F = 31{,}308 \div 396 = 79.06$, which for 1 against 352 degrees of freedom is

Table 11. Comparative Efficiencies of Matched Versus Unmatched Samples of Quarter-Section Grids. 1938-1939 Data.

Item	Number of pairs of unmatched grids equivalent to one matched grid
1. Number of farm acres	18.0
2. Number of corn acres	14.6
3. Number of oat acres	6.6
4. Number of barley acres	3.8
5. Number of swine	5.4
6. Number of horses	10.6
7. Number of cattle	8.3
8. Number of sheep	12.2
9. Number of chickens	9.4
10. Receipts from sales of dairy products	5.8
11. Gross expenditures, operator	3.5
12. Gross income, operator	6.2
13. Net cash income, operator[a]	2.2
14. Number of persons on farm	12.3

[a]Cash grain area not included.

highly significant according to Snedecor's F-table. Hence it seems reasonable to believe that there has been an actual increase of swine during 1938.

The year × county interaction is statistically highly significant ($F = 627 \div 396 = 1.58$) which would suggest that in regard to swine inventories the counties did not hold the same relative positions with one another for the 2 years, indicating that components c and b are probably not independent. This does not, however, affect the con-

clusions reached on the comparative efficiencies, but may have some economic significance.

Year-to-year changes as percentage changes

Often the value of an item in one year is expressed as a percentage of that of the previous year or some other base year. In this case absolute values for either year are of no importance in themselves.

If the amount of an item on a sampling unit enduring through time is x_i in the initial year and y_i for some subsequent year then

$$p_i = \frac{y_i}{x_i} \times 100 \qquad (16)$$

where p_i is the per cent which the subsequent year is of the initial year for the given item. For a population of N sampling units

$$P = \frac{\sum y_i}{\sum x_i} \times 100 \text{ where } i = 1,2 \ldots N \qquad (17)$$

And for a sample of n an estimate of P is given by

$$p = \frac{\sum y_i}{\sum x_i} \times 100 \text{ where } i = 1,2 \ldots n \qquad (18)$$

Roughly, the variance of p is given by

$$\sigma_p^2 = (100)^2 \left(\frac{N-n}{Nn}\right) P^2 \left(\frac{\sigma_y^2}{\bar{Y}^2} + \frac{\sigma_x^2}{\bar{X}^2} - \frac{2\sigma_y\sigma_x\rho_{xy}}{\bar{Y}\bar{X}}\right) \qquad (19)$$

which can be estimated if statistics derived from sample data are available. The square root of this variance provides a rough standard error for the estimated p's.

Estimates of changes from 1938 to 1939 have been computed for a set of items together with their corresponding standard errors, and are shown in the following table accompanied with preliminary and final estimates from AMS publications.

It can be seen from the table that the survey sample was remarkably accurate in estimating changes in the important acreage and livestock items, in fact more accurate than the preliminary estimates of the AMS (if its final estimate is taken as the better of the two). Barley acreage was difficult to measure as shown by the large standard error of its sample estimate. For sheep, chickens, and receipts from dairy products the discrepancy between the sample and AMS estimates appears to be statistically significant. No reason for this is known. For remaining items where comparable data are available agreement for the two sources is apparent.

Errors

Data taken by interview and by sample can usually be rightly suspected of containing error of one sort or another. We are concerned here with the problem of determining the nature and extent of this error.

Errors in data taken by interview

By design, data were collected to test the memory of interviewees for error. The livestock section in the questionnaire used on the 1939

Table 12. Sample Survey Estimates of Per Cent Changes from 1938 to 1939 for a Number of Items Together With Their Standard Errors and Corresponding AMS Preliminary and Final Estimates. Data from 452 Quarter-Section Grids. State of Iowa.

	1939 as per cent of 1938			Standard error of sample survey estimate
	Sample survey	AMS Preliminary[a]	AMS Final[b]	
1. Acres in farms	100.8			1.1
2. Corn acres, all harvested	91.6	93.0	91.3	1.1
3. Oat acres, grain	83.4	85.0	84.0	1.5
4. Barley acres, grain	137.5	126.0	129.1	12.5
5. Number of swine	136.5	118.0	131.0	3.2
6. Number of horses	99.6[c]	96.0	96.0	1.5
7. Number of cattle	108.2	105.0	105.0	2.2
8. Number of sheep	93.7[c]	108.0	104.6	5.0
9. Number of chickens	109.9	102.5	102.5	1.7
10. Receipts from sales of dairy products	104.9[c]	92.2		3.0
11. Gross expenditures, operater	117.0[d]			5.0
12. Gross receipts, operator[e]	123.3[d]	110.9[f]		3.9
13. Net cash income, operator[e]	133.8[d]			7.7
14. Number of persons on farms	102.6[c]			1.3

[a] From preliminary reports.
[b] Crops from December (1940) crop report. Livestock from February (1941) livestock report.
[c] Deviation from 100 not statistically significant.
[d] Not accurate because of changes in questionnaire design. Probably high.
[e] Cash grain area excluded because of insufficient data.
[f] Total agricultural receipts to farms (including landlords' share) hence not directly comparable with survey figure.

survey was constructed in the much used form wherein beginning inventory numbers + numbers raised and bought + change in inventory numbers could be checked on the spot with ending inventory numbers + numbers sold, butchered and died. If discrepancies were detected, adjustments were made in cooperation with the farmer whenever possible. With this kind of statement on the number of the several kinds of livestock on the farm 12 months ago, we had the previous year's statement from the same farmer (on the matched sample) on the numbers he had on hand at that time. Similar data were obtained on feed stocks except that no cross checks were attempted. Farmers were not informed of the test being made on their ability to remember, hence some wondered why we were again asking for information they had previously given us. The results of this test are shown in Tables 13 and 14.

No differences have been detectable among type-of-farming areas. Renters show an inclination to be slightly more inaccurate than owners, which might well be due to the added complexities of rental transactions.

We conclude that the discrepancies shown in Table 13 represent what may be termed memory biases on those items. It is not known how consistent these biases might be through time or how different they might be if questionnaires were of different design. As evidence on the effect of questionnaire design the data in Table 15 may be considered.

It appears that the more complete and detailed questionnaire of the 1939 survey was getting more accurate information than its briefer predecessor. In addition to bias there is a large random component in the errors of memory.

Table 13. Comparison of Inventories (as of Jan. 1, 1939), Reported by 396 Matched Farmers on the Two Dates, Jan. 1, 1938, and Jan. 1, 1939

Item	Remembered as a per cent of previous report
1. Number of cattle, all ages	91
2. Number of swine, all ages	81
3. Number of chickens, all ages	92
4. Bushels of corn, unsealed	92
5. Bushels of oats	84

Table 14. Per cent of Farmers by Tenures Who Have Failed to Remember Accurately Their Inventories of a Year Ago. (For Only Those Having Reported Some Quantity at Either Time.)

Item	Tenure group			
	Owners	Renters	Part-owners	All tenures
	(%)	(%)	(%)	(%)
1. Number of cattle, all ages	76	79	68	76
2. Number of swine, all ages	73	82	58	76
3. Number of chickens, all ages	76	75	80	76
4. Bushels of corn, unsealed	84	90	83	87
5. Bushels of oats	70	78	84	75

Table 15. Sample Survey Estimates of Selected Economic Items as Percentages of the Corresponding AMS Estimates, State of Iowa, 1938 and 1939

Item	1938[a]	1939[b]
	(%)	(%)
Government payments	77	80
Receipts from sales of:		
Cattle, calves	73	84
Hogs	60	84
Sheep, lambs	39	54
Chickens	-	45
Eggs	60	76
Dairy products	74	85

[a]Based on AMS revised estimates.
[b]Based on AMS preliminary estimates.

Discrepancies between reports to the township assessor and the sample survey enumerator

To test for possible discrepancies between the reports given the two data-collecting agencies, sample survey farms were identified in the assessor records and the relevant data compared. This was done for both 1938 and 1939. Of the 773 sample survey farms of 1938 only 576 could be completely identified (those having approximately the same name and within 10 acres of the same size of farm). The remaining 197 were of two kinds: one, 76 which could not be found listed at all and two, 121 found listed but having sizes differing 10 acres or more from the size reported in the survey. Table 16 summarizes the effect this procedure has had on the representativeness of the data.

Table 16. Numbers of Sample Survey Farms Which Could Be Completely Matched, Partially Matched and Not Matched at All (Not Found) in the Assessor Records, 1938, Given by Tenure Group with Mean Farm Size

Class	Owner		Renter		Part-owner		All	
	%	Average size	%	Average size	%	Average size	%	Average size
Total farms (773)	36	148	54	191	10	206	100	177
Completely matched (576)	37	147	56	185	7	188	100	171
Partially matched (121)	33	193	46	228	21	240	100	219
Not found (76)	45	112	46	175	9	155	100	145

We conclude that the group of farms for which reports are available from both sample survey and assessor, are somewhat smaller than the original group. It appears also that there is no significant difference in the proportions among the tenure groups although there is some evidence that part-owners reported quite different farm acres to the two agencies.

However, we believe that this group will be quite useful in an investigation of discrepancies in reports to the two agencies. Table 17 presents a comparison of totals reported by both agencies for a selected list of items.

We see in Table 17 that except for sheep, livestock items are definitely biased. As shown elsewhere (page 17) the sample survey figures agree well with AMS estimates and therefore we conclude that it is the assessor who receives the understatements. Among other items showing a bias is corn

yield. We now have some evidence that difference between the sample survey and assessor corn yields (see Table 2) are real and not likely the result of sampling variation. We present the following data from Table 2:

Year	State census (assessor) (bu./acre)	Sample survey (bu./acre)
1938	46.3	47.7 ±.5
1939	52.2	54.6 ±.4

In both years the sample surveys obtained higher yields. No data are available for determining which is closer to the true yields.

The random errors as measured by the coefficients of variability of the differences are indications of the extent of errors in data taken by interview. These are the results of misunderstandings, vagueness, indifference, deliberate misstatement and to a small extent, errors of memory (the enumerators appeared at different times - sometimes as much as 2 months apart). Except for the last reason both enumerator and enumeratee may be at fault. It may be noted that acres in corn is quite reliably stated (judging from the relatively low random error). Acreage control programs have probably helped to acquaint many farmers with their exact acreage in corn.

Bias which may result from sampling procedure

Enumerators were instructed to visit those farms, the farmsteads of which were situated on the selected quarter-section grids. If information could not be obtained from any of these designated farms, they were instructed to visit the nearest farm as a substitute. Since this was a relaxation of strict sampling procedure, made necessary because we

Table 17. Summary of Bias and Random Error in the Reports of Farmers to the Sample Survey and Iowa Assessor. 1938 and 1939 Data

Item	Total reported to assessor as % of sample survey		Bias[a] (departure of assessor from sample survey) in per cent		"Random" error (coefficient of variability of the differences[b] in per cent of sample survey mean)	
	1938	1939	1938	1939	1938	1939
Corn acres, total	100.4	99.4	-	-0.6	-	6
Corn acres, harvested	100.8	98.9	-	-1.1	-	7
Corn production (bu.)	97.9	97.1	-	-2.9	-	13
Corn yield (bu./acre)	98.1	97.8	-	-2.2	-	10
Oat acres, grain	100.2	100.8	-	-	-	14
Oat production (bu.)	99.8	98.4	-	-	-	20
Oat yield (bu./acre)	101.3	97.5	-	-	-	
Wheat acres, grain	104.8	93.4	-	-	-	24
Wheat production (bu.)	97.9	102.1	-	-	-	42
Wheat yield (bu./acre)	98.0	99.1	-	-	-	
Barley acres	105.0	93.2	-	-	-	26
Barley production (bu.)	106.5	89.4	-	-	-	47
Barley yield (bu./acre)	97.8	89.5	-	-	-	
Alfalfa acres, hay	98.6	93.3	-	-6.7	-	35
Pasture acres, all		100.7		-		25
Horses and mules	96.7	-[c]	-3.3	-	17	-
Cattle	93.9	-[c]	-6.1	-	26	-
Sheep	99.2	-[c]	-	-	88	-
Swine	87.4	-[c]	-12.6	-	27	-
Cows milked during year	91.1	-[c]	-8.9	-	21	-
Sows and gilts bred	97.9	-[c]	-2.1	-	37	-

[a]Differences have been designated a "bias" only when they show statistical significance. Blank spaces indicate that no bias has been detected.

[b]Residual variation after farm differences and bias has been removed. For those reporting "some" to either assessor or sample survey - that is, those reporting "none" to both assessor and survey were excluded in the analysis of random errors.

[c]This information was not obtained by the assessor in 1939.

were dealing with people, we were interested in getting some idea of whether or not this failure to get the original selected farms would result in a biased sample. Consequently enumerators were requested to record the tenure and size of those farms which were not enumerated, together with the reason. Both in 1938 and 1939 it was necessary to substitute 29 per cent of the farms first visited. The number of farms visited but not enumerated, and reasons therefor, are listed in Table 18.

Table 18. Number of Farms for Which Substitutions Were Made Listed by Reason Given by Enumerator, 1939 Survey

Reason	Number of farms	Number in group
I. Operator living on farm		
A. Operator at home		
1. Uncooperative		55
a. Landlord would object	2	
b. Dislike for AAA	2	
c. Dislike government interference	20	
d. Afraid information goes to packers	4	
e. Assessor advised against giving information	1	
f. Resented being singled out for sampling	1	
g. Gave unreasonable data	3	
h. Other, or not given	22	
2. Cooperative, apparently, but		44
a. Busy	29	
b. Sickness	8	
c. Too difficult to reach	5	
d. Drunk	2	
B. Operator not at home		76
a. At a sale, in town	21	
b. Visiting	7	
c. Vacationing	1	
d. No reason given	47	
II. Operator not living on farm		6
A. Absentee operator	3	
B. Nobody on farm at present	3	
III. No reason given	15	15
Total	196	

In Table 19 are summarized the data from the enumerators' reports on the tenure and size of the non-enumerated farms together with the enumerated farms of 1938 and 1939.

Table 19. Number, Per cent and Size of Farm by Tenure Group for the Non-enumerated Farms of 1939 and the Enumerated Farms of 1938 and 1939

Farm group	Owner			Renter			Part-owner			All tenure		
	No.	%	Average size	No.	%	Average size	No.	%	Average size	No.	%	Average size
Non-enumerated, 1939	67	43	163	80	51	166	9	6	210	156	100	167
Enumerated, 1939	292	38	154	398	51	179	88	11	221	778[a]	100	175
Enumerated, 1938	278	36	148	415	54	191	80	10	206	773	100	175

[a]Four managed farm excluded.

We conclude from the data presented in Table 19 that no perceptible bias on either farm size or tenure is evident.

Discussion on errors in data taken by interview

In general, errors due to inaccuracies in the data appear to be larger than errors due to sampling (where the sample is of the size of the two surveys). Except for the unbiased items, further increase in size will scarcely increase the accuracy of sample information. Certainly a complete census does not provide accurate information by the mere fact of complete enumeration.

It has been suggested that improvements in the design of the questionnaire have shown real increases in accuracy. Better education of the enumerators will also help. But there still remains the problem of minimizing errors due to bad memory on the part of the interviewee.

As an experiment, several questionnaires on which beginning inventories from the previous years' record were posted, were tested in the 1939 survey. It seemed the farmers were quite satisfied in having the enumerator remind them of the facts 12 months past. Sales which would have otherwise been overlooked were picked up and any changes in farm population, farm size, feed stocks, etc., were easily detected and checked on the spot. Matching farms without providing the enumerator with all relevant previously obtained information resulted in errors which are quite damaging to matched samples where change is being measured. In our case probably a good part of our "sampling errors" is really variation due to these inaccuracies.

Further lessening of errors of memory can come from shortening the period over which the interviewee is called upon to remember. If data are required over a fiscal year, probably more than two visits will be advisable. Or perhaps some simple account system could be devised by which farmers could be persuaded to record certain transactions without much effort. This could be merely a request that the cooperating farmer keep transaction slips available for the enumerator. Such simple bookkeeping might be offered as a free service for his cooperation. Even with all this, however, recalcitrants will continue to be a problem.

Effect of Stratification (Complete) on Sampling Efficiency

By stratification is generally meant the division of the population under inquiry into two or more parts known as "strata". For instance the population of Iowa farms is "stratified" if it is regarded as composed of owners, renters, part-owners and managers; or as Allamakee County farms, Adams County farms, etc. If two conditions can be met, stratification can improve efficiency of sampling when an accurate estimate of the overall mean is desired. First, strata must be unlike (owners as a group must be different than renters as a group in the character being measured) and second, the total number of elements in each stratum must be known. If these conditions have been satisfied, either one of two usual sampling procedures can be adopted.

If a population is divided into K strata having

$$N_1, N_2, \cdots, N_K$$

sampling units and

$$\sigma_1, \sigma_2, \cdots, \sigma_K$$

standard deviations, the most efficient sample will be composed of

$$n_1, n_2, \cdots, n_K$$

sampling units from the several strata such that

$$\frac{n_1}{N_1\sigma_1} = \frac{n_2}{N_2\sigma_2} = \cdots = \frac{n_K}{N_K\sigma_K} \qquad (20)$$

where $\sigma_1 = \sigma_2 = \cdots \sigma_K$, equation (20) becomes

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \cdots = \frac{n_K}{N_K} \qquad (21)$$

In the usual case, where the σ's are unknown beforehand, stratified samples are allocated according to condition (21), which implies that the σ's have been assumed equal. This is the case of our sample surveys.

We are now in a position to speculate on the merits of both the method of the sample surveys (the choice of the county as the stratum and the assumption of equal σ's) and of alternatives which can be proposed.

The relative efficiencies of stratifications can be obtained directly from the variances within the several kinds of strata. Variances within township, counties and type-of-farming areas are most easily obtained by analysis of variance.[10] In Table 20 are presented efficiencies which may be expected if the survey samples were stratified by township and type-of-farming area or completely unstratified compared with stratification by counties.

It can be seen that there is considerable difference in the way individual items behave but that in general the township is more efficient than the larger stratification units. The type-of-farming area is only slightly less efficient than the county (indicating a relatively high degree of similarity among the counties of which it is composed). With no stratification at all the average loss for the items investigated amounted to 10 per cent both years. For corn yield this loss was a high as 31 per cent and 40 per cent.

We conclude that except for certain individual items, the statistical

---

[10] In our case a correction was theoretically necessary because the survey samples were not random without restriction. See Cochran (1939). It was found, however, that in this case the corrections were so small that they could be ignored.

Table 20. Relative Efficiencies of Samples Stratified by Townships, Type-of-farming Areas and Drawn Without Stratification from the State Compared with Samples Stratified by Counties Such as the Sample Surveys - 1938 and 1939[a] Data

(Figures represent per cent efficiencies. Efficiency of county stratified samples taken as 100.)

Item	1938			1939		
	Town-ships	Areas	State	Town-ships	Areas	State
1. Number of swine	104	100	97	110	84	83
2. Number of horses	105	95	95	183	112	112
3. Number of sheep	54	100	97	97	100	99
4. Number of chickens	103	95	90	90	112	110
5. Number of eggs yesterday	105	95	89	129	89	88
6. Number of cattle	96	97	96	108	99	98
7. Number of cows milked yesterday	78	96	88	74	96	80
8. Number of gallons milked yesterday	80	92	89	88	93	80
9. Receipts from sales of dairy products	78	95	85	-	-	-
10. Number of farm acres	101	101	101	73	96	95
11. Number of corn acres	74	92	80	95	92	79
12. Number of oat acres	66	84	75	105	82	71
13. Corn yield	120	83	69	123	73	60
14. Oat yield	104	91	90	157	92	73
15. Commercial feed expenditures, farm	291	98	95	-	-	-
16. Total cash expenditures, operator	163	97	94	94	103	98
17. Total cash receipts, operator	191[b]	106[b]	104[b]	131	106	101
18. Net cash receipts, operator	148[b]	104[b]	103[b]	113	101	101
19. Number of hogs sold	-	-	-	95	85	84
20. Number of cattle sold	-	-	-	318	103	103
21. Number of hogs bought	-	-	-	810[c]	113	113
22. Number of cattle bought	-	-	-	167	98	97
23. Number of cows and heifers milked during year	-	-	-	74	99	79
Average	115	96	91	121[c]	97	91

[a]Computations on unmatched farms only, therefore independent of the 1938 sample.

[b]Cash-grain area not included.

[c]"Number of hogs bought" not included in average.

gain from geographic stratification is not very large for Iowa except when carried to the township. Since there are about 1600 townships in the state, this means that complete township stratification would require samples of at least 1600 sampling units, and therefore would be feasible only for large samples at best. A decision on relative merits of county versus area stratification is not directly available. It appears that the average loss of 3 per cent or 4 per cent obtained by shifting from the county to the type-of-farming area as the stratum is roughly balanced by savings in cost. The two, therefore, should be approximately equivalent for census-type inquiries. The case for no stratification at all has no appeal mainly because certain items would be estimated with great inefficiency, the savings in cost would not be very much over that where type-of-farming areas are stratified and usually information is desired by type-of-farming areas anyway.

The 1939-survey data was examined to determine the efficiencies available in a stratification based on a farm classification scheme. The classification scheme chosen for this investigation was that proposed by Jebe (1941). Jebe's scheme grouped farms into seven classes designed to bring about the greatest possible degree of homogeneity within classes in regard to eight items. A stratification based on this scheme appears to provide greater sampling efficiency than one based on the usual five type-of-farming areas of Iowa. The relative efficiencies of the two methods of stratification are shown in Table 20a. The figures given in this table represent the estimated efficiency of the farm classification as

Table 20a. Relative Efficiency of a Farm Classification Stratification by Type-of-farming Area, State of Iowa, 1939

Item	Efficiency of farm classification compared with type-of-farming area taking the efficiency of the latter at 100%
	(%)
1. Farm acres	116
2. Corn sales ($)	138
3. Cash operating expenditures ($)	121
4. Cattle sales ($)	137
5. Swine sales ($)	157
6. Dairy products sales ($)	131
7. Cash receipts ($)	136
8. Net income[a] ($)	118

[a]Net cash receipts with adjustment for inventory change.

compared with type-of-farming area stratification where the efficiency of the latter is taken as 100. In every case stratification by the classification scheme is more efficient. In practice, however, stratification by some farm classification scheme would require, if estimates for all farms are desired, relatively accurate information on the relative sizes of the classes (strata). For similar reasons the sizes of the type-of-farming areas must be known with reasonable accuracy. In the case of the sample surveys the sizes of the type-of-farm classes were not known. We conclude, therefore, that until additional information is obtained on the relative sizes of farm classes, the type-of-farming area (and other geographic strata) is the recommended basis for stratification.

The possible merits of stratification by tenure group (owner, renter and part-owner) were investigated by means of analysis of variance on a selected group of items. In Table 21 are presented item means by tenure, tests of the significance of their differences and the efficiency of a sample stratified by tenure compared with one drawn at random in the state.

Table 21. Item Sample Means by Tenure and Relative Efficiency of Tenure Stratification Compared with no Stratification, State of Iowa, 1939

Item	State mean per farm by tenure, 1939				Relative efficiency
	Owners	Rent-ers	Part-owners	All	
					(%)
1. Number of swine	51.2	51.6	62.0	52.6	101.0
2. Number of horses and mules	3.74	4.12	4.81	4.06	101.4*
3. Number of sheep	4.20	5.35	12.2	5.68	100.0
4. Number of chickens	164.	165.	153.	163.	99.6
5. Number of eggs yesterday	25.4	22.9	22.3	23.8	99.6*
6. Number of cattle	25.4	22.9	30.1	24.3	101.6
7. Number of cows milked yesterday	4.62	4.94	6.02	4.95	100.6
8. Number of gallons milked yesterday	9.93	10.57	11.70	10.5	99.7
9. Number of farm acres	157.	179.	221.	175.	103.0**
10. Number of corn acres	38.3	51.3	62.4	47.7	105.8**
11. Number of oat acres	19.3	28.1	29.1	24.9	102.0*
12. Corn yield per acre	57.3	53.4	52.9	54.5	102.1*
13. Oat yield per acre	32.0	30.8	30.4	31.0	101.2
14. Net income, operator ($)	1252.	928.	1607.	1128.	103.8
15. Number of hogs sold	43.5	42.2	47.9	43.4	99.6
16. Number of cattle sold	12.5	7.7	15.1	10.4	100.7
17. Number of hogs bought	3.9	4.6	3.1	4.2	99.3
18. Number of cattle bought	8.3	4.8	11.2	6.9	100.3
19. Number of cows milked during year	7.0	7.1	7.9	7.1	99.7
Average					101.1

*Statistically significant at 5 per cent level.
**Statistically significant at 1 per cent level.

Table 22. Items Having Large Differences Between Tenure Groups, 1939 Survey Data

Item	Mean per operator			
	Owner	Renter	Part-owner	All tenure
Number of persons on farms, 1/1/40, farm	4.024	4.530	5.034	4.405
Number of persons born during 1939, farm	.0445	.1030	.0455	.0742
Number of persons died during 1939, farm	.0240	.0352	.0455	.0320
Receipts from machine work, 1939, operator	18.42	43.91	49.23	34.86
Receipts from labor, non-farm, 1939, operator	40.59	9.44	12.88	21.67
Receipts from labor, farm, 1939, operator	2.40	11.89	1.31	7.11
Receipts from "other income"[a], 1939, operator	42.74	20.93	28.92	30.08

*Includes pensions, income from sales work, etc.

We note that for the items shown there are few having very large tenure differences, and any gain in efficiency by tenure stratification is almost negligible.

There are items, however, where tenure differences are large. As an example a few have been selected from survey data and are shown in Table 22.

We conclude that except for some special inquiries, stratification by tenure does not promise to be very effective. Furthermore, there still remains the problem of determining the sizes of these tenure groups before tenure stratification can be used.

Efficiency in the allocation of the sampling units between and within counties: incomplete stratification or subsampling

We wish here to determine the effects on sampling efficiency resulting from different geographical allocations of the quarter-sections selected for the sample. For example, what efficiency would we expect from the same 900 quarter-sections if, instead of having 9 selected from each of 100 counties, 18 were selected from 50 counties (both quarters and counties taken at random). We might also wish to know if sampling

Table 23. Analysis of Variance of Number of Cattle per Farm, State of Iowa, 1938

Source of variation	Degrees of freedom	Sum of squares	Mean square
Total	772	382,185	
Type-of-farming area	4	3,708	927.0
Counties within areas	96	59,345	618.2
Townships within counties	421	209,527	497.7
Quarters within townships	101	52,857	523.3
Farms within quarters	150	56,748	378.3

efficiency could be improved through the use of various proposed stratification procedures.

To answer these and related questions, we again find it convenient to use analysis of variance procedure. For this, a typical analysis of variance is summarized in Table 23.

We note in the table that excepting the township each geographical

unit seems to be contributing variation to the population of farms. This being the case quarters and township mean squares for this item can be pooled, giving the following analysis of variance.

Source of variation	Degrees of freedom	Mean square
Total	772	
Area	4	927.0
Counties within areas	96	618.2
Quarters within counties	522	502.6
Farms within quarters	150	378.3

Variance of the sample estimate of mean number of cattle where the sample is taken in the manner of the 1938 survey will be given by mean square of quarters within counties divided by total number of farms or

$$\sigma^2_{\bar{z}} = \frac{502.6}{773} = .6502$$

The standard error will be $\sqrt{.6502}$ or .81 head.

If number of quarter-sections were doubled within each county, variance of the sample mean, or $\sigma^2_{\bar{z}}$, would be halved (approximately, since the number of farms so selected would not necessarily be exactly doubled).

Now if the number of quarters within counties (sampled) were doubled but the number of counties sampled halved, then $\sigma^2_{\bar{z}}$ will be given by the formula[11]

$$\frac{1}{773C}\left(A(C-c) + Bc\right) \qquad (22)$$

---

[11] This has been derived from the general formula for incomplete stratification of finite populations:

$$\frac{A}{k}\left(\frac{1}{c}-\frac{1}{C}\right) + \frac{B}{C}\left(\frac{1}{k}-\frac{1}{K}\right) \qquad (22.1)$$

where k and K are the number of farms per county in the sample and population, respectively. Since K is large (about 2000 farms) then

where C = total number of counties in an area (= 20.2)

c = number of counties in each area selected for sampling (= 10)

A = mean square between counties within type-of-farming area

B = mean square between quarters within counties

$$\frac{1}{773}\left(618.2\ (20.2-10) + 502.6 \times 10\right)$$

$$= .7257$$

The relative efficiency of this method with respect to the first will be

$$\frac{.6502}{.7259} \times 100 \text{ or } 89 \text{ per cent, a loss of } 11 \text{ per cent.}$$

Computations have been carried through in a similar manner for a group of different items, which are summarized in Table 24.

---

[11](Continued)
(22.1) becomes

$$\frac{1}{Cck}\left(A\ (C-c) + Bc\right) \qquad (22.2)$$

In our case where we are dealing with 5 type-of-farming areas, therefore

$$\sigma^2_{\bar{z}} = \left(\frac{1}{5}\right)\frac{1}{Cck}\left(A\ (C-c) + Bc\right) \qquad (22.3)$$

but since 5ck = 773 farms,

$$\sigma^2_{\bar{z}} = \frac{1}{773C}\left(A\ (C-c) + Bc\right)$$

This formula will apply only approximately to our case where the number of counties vary by area, the number of quarter-sections vary by county and the number of farms vary by quarter-section. The approximation, however, should be adequate for the purpose at hand.

We see that for the items investigated the resulting loss in efficiency would have been on the average 10 per cent in 1938 and 5 per cent in 1939, the greatest loss being 39 per cent for oat acreage in 1938. Apparently no loss would have been made in some items such as "total cash receipts" and "net cash income". (Since these efficiencies were based on sample data they are therefore subject to sampling variation. Consequently individual efficiencies are to be taken with caution.)

On the cost side it seems likely that such a sampling scheme would reduce costs within county about 11 per cent or overall costs at least 5 per cent.[12]

This being the case it appears that both sample schemes would provide on the average about the same amount of information for the money spent.

In general, if fairly good estimates are desired on each of a wide range of items it appears that sampling counties (that is, taking only a fraction of the counties into the sample) is not advisable. For income estimates alone it seems that sampling counties would be quite advisable under the 1938 and 1939 circumstances.

Another argument for sampling counties is that concentrating the areas worked permits greater control over the field crew. When complicated questionnaires (such as those designed to obtain income) are used it may be advisable to have supervisors meet frequently with enumerators during

---

[12]Based on sample survey conditions: a one-hour questionnaire, $8.00 a day salary and expense and 5¢ per mile. See pp. 72 to 81 for details on cost.

Table 24. Estimated Relative Efficiency (Compared with the Sample Surveys) by Which Selected Items Would Be Sampled if Sampling Were Doubled Within Counties and the Number of Counties Halved. Counties Stratified by Type-of-farming Area. 1938 and 1939[a] Data

Item	Relative sampling efficiency (1938 and 1939 surveys = 100)	
	1938 (%)	1939 (%)
1. Number of swine	99	76
2. Number of horses	85	108
3. Number of sheep	98	100
4. Number of chickens	85	122
5. Number of eggs yesterday	87	84
6. Number of cattle	89	99
7. Number of cows milked yesterday	88	93
8. Number of gallons milked yesterday	77	89
9. Receipts from sales of dairy products	87	-
10. Number of farm acres	103	94
11. Number of corn acres	76	88
12. Number of oat acres	61	74
13. Corn yield per acre	88	89
14. Oat yield per acre	77	65
15. Commercial feed expenditures, farm	92	-
16. Total cash expenditures, operator	91	104
17. Total cash receipts, operator	119[b]	109
18. Net cash income, operator	113[b]	102
19. Number of hogs sold	-	76
20. Number of cattle sold	-	104
21. Number of hogs bought	-	123
22. Number of cattle bought	-	97
23. Number of cows and heifers milked during year	-	99
Average	90	95

[a] Computations on unmatched farms only, therefore independent of the 1938 sample.

[b] Cash-grain area not included.

the survey. Other savings may be made depending, of course, on circumstances of the survey (whether or not photographic maps in county AAA offices are consulted).

We conclude that for census-type questionnaires (where a variety of items are asked for information on each) sampling counties is not advisable in Iowa. For an income survey, however, it seems that this procedure is advisable for years which are not too unlike 1938 and 1939. If there is uncertainty, the all-county plan is recommended.

Homogeneity of variances

It was stated previously that the number of sample quarter-sections drawn from each county was made proportional to the total number of quarter-sections in that county. This was done because knowledge of the variances within counties or other strata was not available. However, we now have data which will provide estimates of the variances within some of these strata and therefore we can determine how good or how bad this and other alternative sampling allocations are or would be. We can now compare the relative efficiencies of stratified samples allocated as

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \cdots = \frac{n_K}{N_K} \qquad (23)$$

with samples allocated as

$$\frac{n_1}{N_1\sigma_1} = \frac{n_2}{N_2\sigma_2} = \cdots = \frac{n_K}{N_K\sigma_K} \qquad (24)$$

Table 25. Estimated Variances of Quarter Sections (on a Farm Basis) in Each Type-of-Farming Area and for the State, for a Selected List of Items, 1938 and 1939

Item	Year	Northeast dairy	Cash grain	Western livestock	Southern pasture	Eastern livestock	State
1. Number of swine	1938	984	1,584	1,795	1,355	2,563	1,708
	1939	1,926	2,352	2,767	1,967	2,235	2,303
2. Number of horses and mules	1938	7.98	9.22	9.58	8.70	7.40	8.63
	1939	5.75	7.73	8.03	6.27	5.39	6.66
3. Number of sheep	1938	127	1,698	173	1,003	301	690
	1939	764	20	618	209	87	235
4. Number of chickens	1938	12,212	12,090	7,370	6,858	6,934	9,600
	1939	15,426	14,467	12,663	10,713	5,181	12,043
5. Number of eggs yesterday	1938	665	620	1,028	432	298	761
	1939	1,236	690	710	432	412	721
6. Number of cattle	1938	382	402	514	522	790	523
	1939	312	480	384	306	255	356
7. Number of cows milked	1938	20.9	11.5	10.9	7.7	13.9	14.1
	1939	23.6	23.5	4.9	6.8	13.3	17.7
8. Number of gallons milked	1938	137.2	96.8	49.2	42.7	95.2	86.1
	1939	179.7	93.1	48.5	21.1	77.3	99.9
9. Receipts from dairy products	1938	119,912	127,570	60,400	29,817	78,007	94,596
	-	-	-	-	-	-	-
10. Number of farm acres	1938	8,403	9,929	12,148	21,666	11,728	12,669
	1939	5,559	8,345	15,696	17,052	6,788	10,572
11. Number of corn acres	1938	824	1,668	2,388	1,262	1,069	1,716
	1939	446	993	1,262	946	803	1,043
12. Number of oat acres	1938	587	1,008	1,585	451	396	949
	1939	390	771	396	392	265	521
13. Corn yield	1938	194	155	178	172	264	266
	1939	236	171	193	172	214	251
14. Oat yield	1938	123	106	155	105	217	143
	1939	122	160	209	78	97	167
15. Commercial feed expenditures	1938	10,833	18,688	10,144	24,511	159,917	43,650
	1939	-	-	-	-	-	-
16. Total cash expenditures, operator	1938	899[a]	3,683[a]	1,634[a]	624[a]	5,539[a]	2,597[a]
	1939	1,001[a]	4,536[a]	4,570[a]	2,496[a]	2,387[a]	3,149[a]
17. Total cash receipts, operator	1938	3,015[a]	-	3,770[a]	2,429[a]	9,961[a]	4,934[a]
	1939	2,155[a]	6,030[a]	7,925[a]	2,044[a]	4,770[a]	4,920[a]
18. Net cash income, operator	1938	1,148[a]	-	1,406[a]	989[a]	1,585[a]	1,315[a]
	1939	933[a]	2,425[a]	2,633[a]	612[a]	1,784[a]	1,729[a]
19. Number of hogs sold, farm	1938	-	-	-	-	-	-
	1939	1,530	1,814	2,988	1,825	1,658	1,973
20. Number of cattle sold, farm	1938	-	-	-	-	-	-
	1939	120	1,991	445	255	232	631
21. Number of hogs bought, farm	1938	-	-	-	-	-	-
	1939	12	80	1,113	84	247	326
22. Number of cattle bought, farm	1938	-	-	-	-	-	-
	1939	56	2,132	565	355	68	671
23. Number of cows and heifers milked	1938	-	-	-	-	-	-
	1939	41.3	23.1	10.9	11.5	38.8	31.7

[a](000) omitted.

For convenience and in order to assure fairly good estimates of the σ's a stratification by type-of-farming area will be considered. Again the analyses of variance provide the necessary information. When corrected for county stratification, the mean square for quarter-sections within a type-of-farming area is an estimate of the σ^2 for that stratum. In Table 25 are shown the sample estimates of the population variances for each of the five areas and for the state as a whole (unstratified) for a selected list of items.

It can be seen that the areas do not have the same relative positions (with one another) in regard to variance. No area is consistently high or low for all items. There is even a tendency to shift relative positions from one year to another on the same item (see number of cattle). Allowance should be made for sampling variation, since these figures are merely estimates of the true variances. It is interesting to note, however, that for this set of items, the Northeast Dairy and Southern Pasture Areas, occur more frequently with lowest variances whereas Western Livestock and Eastern Livestock are found with highest variances. In general, however, there seems little reason for saying that a certain area is more variable than another without regard for the specific items under consideration.

Let us say, however, that we are interested in one item in particular, then what (if any) gain is to be obtained by different allocation? For example let us select an item that appears to have large differences in variances among the areas such as "net cash income to the operator".

We have the following information of the type-of-farming area populations and of the sample (1939).

Type-of-farming area	Number of rural farms		
	Population[a]	Sample	Estimated n
Northeast dairy	39,574	153	121
Cash grain	38,412	163	183
Western livestock	44,017	162	218
Southern pasture	36,935	141	88
Eastern livestock	41,832	163	172
State	200,770	782	782

[a]Derived from Iowa Farm Census data.

If the 782 rural farms drawn for the sample were so allocated among the five areas that

$$\frac{n_1}{N_1\sigma_1} = \frac{n_2}{N_2\sigma_2} = \frac{n_3}{N_3\sigma_3} = \frac{n_4}{N_4\sigma_4} = \frac{n_5}{N_5\sigma_5}$$

then we should use the values of n shown in the above table as "Estimated n."

The best estimate of "net cash income to the operator" (mean per farm for all farms in the state) would be the weighted mean

$$\bar{x}_w = \frac{N_1\bar{x}_1 + N_2\bar{x}_2 + N_3\bar{x}_3 + N_4\bar{x}_4 + N_5\bar{x}_5}{N_1 + N_2 + N_3 + N_4 + N_5} \qquad (25)$$

where $\bar{x}_1$ is the mean for farms in area 1, etc.

N_1 is the total number of rural farms in area 1, etc., and its variance would be estimated by

$$s^2_{\bar{x}_w} = \frac{\Sigma N_i^2 \, s_i^2/n_i}{(\Sigma N_i)^2} \qquad (26)$$

$$i = 1,2,\ldots,K$$

For "net cash income to the operator," $s^2_{\bar{x}_w} = 2{,}079$. This is to be compared with the $s^2_{\bar{x}}$ which would have been obtained if the sample had been drawn at random from each type-of-farming area such that the number of sample farms was proportional to total number of farms in each area, ignoring differences in the σ's. In this case $s^2_{\bar{x}}$ is obtained directly from the analysis of variance as the mean square of quarter-sections within areas for the state divided by 782. It will be found that $s^2_{\bar{x}}$ is about 2,214. The relative efficiency of the two kinds of samplings is

$$\frac{s^2_{\bar{x}}}{s^2_{\bar{x}_w}} = \frac{2214}{2079} = 106\%$$

which indicates that about 6 per cent can be gained for this item by considering variances when allocating the sample within the type-of-farming areas. The gain is not large and what is more, it is not a clear gain since estimates of σ's were used. Moreover, by allocating the sample in this manner some damage has been done to the accuracy of estimating other items of the survey. For example, corn yield in 1939 would have suffered a loss of 7 per cent in sampling efficiency.

From an inspection of the variances of individual items it appears then that no great gains could have been achieved through reallocation of the sample.

The Problem of Maximizing Amount of Information Obtainable From a Given Expenditure by Varying Size of the Sampling Unit and the Number Taken

Up to now where relative efficiencies of alternative sampling schemes were being compared, we have been usually satisfied with making comparisons on the basis of statistical sampling efficiency alone. We shall attempt here to investigate the more practical and also more difficult problem of deciding which sampling schemes provide the most information for the money available.

For simplicity, the case to be considered here will be samples, of which sampling units are of varying size, taken at random within the State of Iowa.

We have the two factors:

y, the number of sampling units taken
and x, the number of farms per sampling unit

which can be varied independently at will by the sampler. Now both sampling variance and cost are functions of these two factors:

$$\text{Sampling variance, } \sigma_{\bar{x}}^2 = f(x,y) \qquad (27)$$

$$\text{Sampling cost, } E = \phi(x,y) \qquad (28)$$

Out objective is to determine what values of x and y will minimize $\sigma_{\bar{x}}^2$ for a given E. To do this we must first determine, if we can, the explicit forms of $f(x,y)$ and $\phi(x,y)$.

A variance function

If we regard the State of Iowa as composed of Y grids of X farms each, then we can set up an analysis of variance (on a farm basis) as follows:

Source	Degrees of freedom	Mean square	Sum of squares
Total	XY - 1	K	(XY - 1)K
Grids	Y - 1	A	(XY - 1)K - Y(X - 1)B
Farms within grids	Y(X-1)	B	Y(X - 1)B

From the table we can write for the grid mean square,

$$A = \frac{(XY - 1)K - Y(X - 1)B}{Y - 1} \tag{29}$$

Suppose now that a sample of y grids was taken, then the variance of sample mean per farm, $\bar{z}$ is given by

$$\sigma^2_{\bar{z}} = \frac{A}{Xy} = \frac{(YX - 1)K - Y(X - 1)B}{Xy(Y - 1)} \tag{30}$$

which becomes when Y is large (that is, when grids are relatively small),

$$\sigma^2_{\bar{z}} = \frac{K}{y} - \frac{(X - 1)}{Xy} B, \tag{31}$$

and in the usual case X is not known but must be estimated from the sample, then $\sigma^2_{\bar{z}}$ must be estimated by

$$s^2_{\bar{z}} = \frac{K}{y} - \frac{(x - 1)}{xy} B \tag{32}$$

Now as a matter of fact for a given grid size the numbers of farms vary from grid to grid and where the grid becomes relatively small (a section or less) some grids will contain no farms at all. Since the

number of degrees of freedom associated with the grid mean square depends on the number of grids having farms, it will be necessary to regard Y and y as the population and sample number of grids having farms, and X and x as the population and sample mean number of farms per grid having farms. (About two-thirds of quarter-sections and about 99 per cent of sections have farms.)

An estimate of K can be obtained from a sample; moreover K is independent of x and y and is therefore a constant. B, the variance of farms within grids, may or may not be independent of x, although it is independent of y. What can we say of the relationship of B and x? Our answer is essentially empirical.

Estimates of B's for the quarter-section, township, county and state are available from the analyses of variance (after proper corrections are made). See Cochran (1939). If the logarithms of these B's are plotted against the logarithms of the corresponding quarter-section, township, county and state areas, it will be seen that for a good number of items a fairly good linear relationship exists. Smith (1938) found that a similar empirical relationship existed between the variances of crop yields and plot areas.

Since it seems to be somewhat more reasonable, mean square distance among points within grids rather than area will be used as a measure of grid size in this empirical function. Hence we can write

$$\log B = \log c_1 + g \log d \qquad (33)$$

or

$$B = c_1 d^g \qquad (34)$$

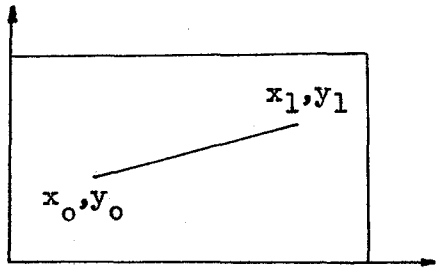
where: c_1 and g are constants (g is the slope of B on d when graphed on double log paper).
d is the mean square distance among points within the grid.

If a and b are the sides of a rectangle[13], $d = \frac{a^2 + b^2}{6}$. When $g = 0$, then $B = c_1 = K$, which would mean that the item concerned is as variable in small groups as in large - that there is no intra-class correlation. If, for convenience, we limit ourselves to the case

---

[13]No reference can be given for this formula in the literature. With the aid of Dr. C.P. Winsor the formula was developed in the following manner:

In the attending diagram the distance between any two points, x_0,y_0 and x_1,y_1 in the rectangle is given by



$$D = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}$$

$$D^2 = (x_1 - x_0)^2 + (y_1 - y_0)^2$$

mean D^2 for all points is given by

$$d = \frac{1}{a^2b^2} \int_0^b \int_0^b \int_0^a \int_0^a (x_1^2 - 2x_1x_0 + x_0^2 + y_1^2 - 2y_1y_0 + y_0^2)dx_1dx_0dy_1dy_0$$

solving,

$$d = \frac{a^2 + b^2}{6}$$

In the case of a square, $b = a$, and therefore $d = \frac{a^2}{3}$

where grids are square, then (34) can be written in terms of x, thus[14]

$$B = c_2 x^{g/2} \qquad (35)$$

We can now write (32) in terms of x, y and determinable quantities, where now

$$s_z^2 = \frac{K}{y} \quad \frac{(x-1)}{y} \; c_2 x^{g-1} \qquad (36)$$

This then is our function, f(x,y), the variance function. Let us now look at cost.

A cost function for sample surveys

If a route connecting y points located at random in a fixed area is minimized, the total distance, D, of that route is[15]

$$D = d\left(\frac{y-1}{\sqrt{y}}\right) \qquad (37)$$

where d is a constant.

---

[14]The intermediate steps are as follows: If a is the side of a square area and k is the number of farms per unit area, then

$$x = ka^2 \qquad \text{and } a^2 = \frac{x}{k}$$

Since for a square, $d = \frac{a^2}{3}$ and since $a^2 = \frac{x}{k}$

then $d = \frac{x}{3k}$

Now $B = c_1 d^g$ can be written

$$B = c_1 \left(\frac{x}{3k}\right)^g$$

or $B = c_2 x^g$ where $c_2 = \frac{c_1}{(3k)^g}$

[15]Found stated in Mahalanobis (1940).

This relationship is based upon the assumption that points are connected by direct routes. In Iowa the road system is a quite regular network of mile square mesh. There are very few diagonal roads, therefore, routes between points resemble those taken on a checkerboard. A test wherein several sets of different members of points were located at random on an Iowa county road map, and the minimum distance of travel from a given point on the border of the county through all the points and to an end point (the county border nearest the last point on route), revealed that

$$D = d\sqrt{y} \qquad (38)$$

works well. Here y is the number of randomized points (border points not included). This is of great aid in setting up a cost function.

To proceed, let:

x = number of farms in a sampling unit

q = time (in hours) spent on a farm. (This covers total time elapsing during the farm visit.)

w = salary and living expenses (in dollars per hour while working)

t = average distance between farms within the sampling unit (in miles)

m = cost per mile of travel (in dollars)

s = average speed of travel (miles per hour)

y = number of sampling units in the sample

then the costs at and among y sampling units will be:

Cost due to	Costs at y sampling units	Costs among y sampling units
Time: enumerating	$yxqw$	
traveling	$yxtw/s$	$\sqrt{ydw/s}$
Transportation:	$yxtm$	$\sqrt{ydm}$

Total cost[16] E, therefore is the sum of these costs or

$$E = yxtm + \sqrt{ydm} + yxqw + yxtw/s + \sqrt{ydw/s}$$

$$= xy(tm + qw + tw/x) + d(m + w/s)y^{\frac{1}{2}}$$

putting $(tm + qw + tw/s) = A$

and $d(m + w/s) = B$

then $E = Axy + By^{\frac{1}{2}}$ (39)

In Iowa, t is approximately a constant having the value 1 mile for points randomly selected within the state while d is roughly 232. The remaining variables will depend on the circumstances of the proposed survey.

We now have an expression for $\phi(x,y)$ the cost function.

With $s_z^2 = \frac{K}{y} - \frac{(x-1)c_2x^{g-1}}{y}$

and $E = Axy + By^{\frac{1}{2}}$,

s_z^2 can be minimized[17] for a given E.

---

[16] Not to be confused with total cost of survey. Only those costs largely affected by x and y have been considered in this cost function.

[17] The minimum can be obtained by minimizing

$$f(x,y) - \lambda\phi(x,y)$$

which gives two equations, $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$; and a third, $\phi(x,y) = E$ to determine the three unknowns: x,y and λ. The first two equations are:

$$\frac{\partial f}{\partial x} = \lambda\frac{\partial \phi}{\partial x}$$

$$\frac{\partial f}{\partial y} = \lambda\frac{\partial \phi}{\partial y}$$

which become, when λ is eliminated,

$$\frac{\partial f}{\partial x}\frac{\partial \phi}{\partial y} - \frac{\partial f}{\partial y}\frac{\partial \phi}{\partial x} = 0$$

Since an algebraic solution of these equations is rather difficult we shall have to adopt a rougher but more convenient procedure of determining the best allocation of expenditure by trial and error.

For investigation we selected seven sampling units, the individual farm and the following six grids: quarter-section, half-section, full section, 2 adjacent sections, 4-section block and the 36-section block (survey township). Assuming (in the cost equation) an s of 30 miles per hour, w to be a dollar per hour and q and m given specified values, total number of sampling units which can be covered for a given expenditure have been computed.[18] They appear in Table 26.

---

[17](Continued)
and when substitutions are made we obtain

$$A\left(K-c_2x^g + c_2x^{g-1}\right) + \left(Ax + \frac{By^{\frac{1}{2}}}{2}\right)\left(-c_2gx^{g-1} + c_2(g-1)x^{g-2}\right) = 0$$

which reduces to

$$y = \frac{B^2}{4A^2x^2}\left\{\frac{g(x-1)+1}{\frac{Kx}{c_2}^{1-g} - x(g+1) + g}\right\}^2$$

This equation together with $E = Axy + By^{\frac{1}{2}}$

provides two equations to determine x and y for the minimum. It can be seen that the complete solution becomes rather difficult.

[18]In addition it was assumed that costs per farm became constant for sampling units exceeding in size the 4-section block. This decision was made because it was felt that where the enumerator must travel to a town for overnight lodging it was no less costly for him to locate himself at a new sampling unit than return to that being worked the previous day. It can be seen that this holds only roughly but it is believed to be a good approximation for the purposes at hand.

Table 26. Numbers of Sampling Units Which Can Be Covered, Given Several Cost Situations, Two Expenditure Levels, and Seven Different Sampling Units,[a] Unstratified Sample in the State of Iowa

Sampling unit	Number of farms: Sampling unit[b]	Mileage at 2¢ / mile: Length of farm visit: 15 min.	60 min.	120 min.	Mileage at 5¢ / mile: Length of farm visit: 15 min.	60 min.	120 min.
A. Total expenditure of $1000							
Individual farm	1.000	1644	650	371	1088	517	315
Quarter-section	0.914	1745	699	401	1140	551	339
Half-section	1.828	1073	392	218	764	336	192
Section	3.656	624	213	116	475	186	105
Two-sections	7.312	347	113	60	278	102	56
Four-sections	14.624	187	59	31	156	54	29
Thirty-six sections	131.616	21	7	4	17	6	3
B. Total expenditure of $2000							
Individual farm	1.000	4012	1452	803	2886	1223	712
Quarter-section	0.914	4293	1569	871	3057	1314	769
Half-section	1.828	2494	852	462	1900	744	421
Section	3.656	1388	451	241	1112	407	225
Two-sections	7.312	749	235	124	623	217	118
Four-sections	14.624	396	121	63	338	113	61
Thirty-six sections	131.616	44	14	7	38	13	7

[a]Computed from the formula: $y = \left(\frac{-B \sqrt{B^2 + 4ACx}}{2Ax}\right)^2$

[b]Computed from the sample survey data.

In Table 27 are shown computed numbers of farms which can be visited for a given expenditure and the corresponding average cost per farm.

Table 27. Estimated Number of Farms Which Can Be Enumerated and Average Cost per Farm, Given Several Cost Situations, Two Expenditure Levels, and Seven Different Sampling Units

Expenditure and Sampling unit	Mileage at 2¢ / mile: Length of farm visit: 15 min. No. of farms	15 min. Cost per farm	60 min. No. of farms	60 min. Cost per farm	120 min. No. of farms	120 min. Cost per farm	Mileage at 5¢ / mile: Length of farm visit: 15 min. No. of farms	15 min. Cost per farm	60 min. No. of farms	60 min. Cost per farm	120 min. No. of farms	120 min. Cost per farm
A. Total expenditure of $1000												
I.F.	1644	$0.61	650	$1.54	371	$2.70	1088	$0.92	517	$1.93	315	$3.17
S_4	1595	0.63	639	1.56	366	2.73	1042	0.96	504	1.99	309	3.24
S_2	1962	0.51	717	1.39	398	2.51	1397	0.72	614	1.63	352	2.84
S	2280	0.44	778	1.29	422	2.37	1737	0.58	680	1.47	385	2.60
2-S	2538	0.39	825	1.21	440	2.27	2034	0.49	744	1.34	411	2.43
4-S	2739	0.37	860	1.16	453	2.21	2277	0.44	791	1.26	430	2.33
36-S	2739	0.37	860	1.16	453	2.21	2277	0.44	791	1.26	430	2.33
B. Total expenditure of $2000												
I.F.	4012	0.50	1452	1.38	803	2.49	2886	0.69	1223	1.64	712	2.81
S_4	3923	0.51	1434	1.39	796	2.51	2794	0.72	1201	1.67	703	2.84
S_2	4559	0.44	1557	1.28	845	2.37	3473	0.58	1360	1.47	770	2.60
S	5076	0.42	1650	1.21	881	2.27	3955	0.49	1447	1.34	799	2.44
2-S	5479	0.37	1720	1.16	907	2.21	4553	0.44	1585	1.26	859	2.33
4-S	5784	0.35	1771	1.13	926	2.16	4936	0.41	1657	1.21	888	2.25
36-S	5784	0.35	1771	1.13	926	2.16	4936	0.41	1657	1.21	888	2.25

B-values were computed for a set of items including both 1938 and 1939 data, where $B = c_1 d^g$. Then with equation (32)

$$v_{\bar{z}} = \frac{100}{\bar{z}} \sqrt{\frac{K}{y} - \frac{(x-1)}{xy}} B \qquad (40)$$

as a formula by which the relative sampling errors of the various sampling units and cost conditions can be computed. A set of these computations appears in Table 28.

In Table 29 the effect of cost factors on overall sampling efficiency of the six-grid sampling units is clearly shown. Low mileage costs, long questionnaires and large total expenditure require smaller grids; and conversely, high mileage costs, short questionnaires and small total expenditure require larger grids.

For a sample survey on the expenditure level of the 1938 and 1939 Iowa surveys (Case V: $1000, 5¢ per mile and 60 minute questionnaire) it looks as if both the quarter- and half-section grid would have about the same efficiency. For certain administrative reasons (not-at-home farms could be revisited more cheaply and conveniently, an accurate determination of the number of farms on the chosen areas can be made more conveniently in the county AAA offices, etc.), the half-section may be recommended over the quarter-section as a sampling unit for this kind of survey. If, however, a much larger sample is taken (greater expenditure) then the quarter-section becomes the better choice. (Compare cases V and XI, Table 29.)

It must be remembered that these observations on the efficiencies of sampling units apply only to the case where item means per farm are being estimated. Efficiencies may be quite different in the case where item means are being estimated on a per grid basis.

Table 28.1. Relative Standard Errors (Per cent of Item Means per Farm) Estimated for Samples of Different Sampling Units and Taken at Random Within the State, 1938 and 1939

(Case I: Expenditure of $1000, 15-minute questionnaire and 2¢ per mile)

Items	Sampling unit						
	I.F.	S_4	S_2	S	2-S	4-S	36-S
	1938						
1. Number swine	2.67	2.82	2.74	2.90	3.36	4.11	9.99
2. Number horses	1.83	1.93	1.87	1.98	2.27	2.80	6.87
3. Number sheep	9.61	9.76	8.80	8.16	7.74	7.44	7.44
4. Number chickens	1.61	1.70	1.66	1.78	2.07	2.57	6.34
5. Number eggs yesterday	3.17	3.21	2.90	2.69	2.55	2.45	2.45
6. Number cattle	2.55	2.67	2.55	2.65	2.98	3.62	8.66
7. Number cows milked	1.98	2.07	2.00	2.09	2.37	2.88	6.79
8. Number gallons milked	2.34	2.45	2.32	2.39	2.64	3.15	7.17
9. Dairy product receipts	2.99	3.11	2.93	2.97	3.24	3.79	8.55
10. Number farm acres	1.54	1.63	1.57	1.64	1.87	2.28	5.58
11. Number corn acres	1.95	2.06	1.98	2.08	2.37	2.87	6.88
12. Number oat acres	2.36	2.59	2.66	3.05	3.78	4.91	12.76
13. Corn yield	.82	.90	.94	1.09	1.36	1.78	4.73
14. Oat yield	.84	.88	.84	.86	.96	1.15	2.71
15. Commercial feed expenditures	6.23	7.06	7.60	9.14	11.78	15.71	43.07
16. Total expenditures, operator	3.96	4.36	4.51	5.21	6.48	8.46	22.36
17. Total receipts, operator	3.16	3.49	3.64	4.23	5.29	6.93	18.39
18. Net cash income, operator	3.54	3.82	3.84	4.26	5.13	6.57	16.82

Table 28.1. (Continued)

1939

Items	Sampling unit						
	I.F.	S_4	S_2	S	2-S	4-S	36-S
1. Number swine	2.16	2.33	2.33	2.58	3.09	3.92	10.01
2. Number horses	1.59	1.62	1.46	1.35	1.28	1.24	1.24
3. Number sheep	6.51	6.82	6.53	6.79	7.64	9.17	21.65
4. Number chickens	1.68	1.75	1.66	1.70	1.86	2.21	4.82
5. Number eggs yesterday	2.73	2.88	2.78	2.93	3.35	4.10	9.98
6. Number cattle	1.98	2.01	1.81	1.68	1.59	1.53	1.53
7. Number cows milked	2.05	2.12	1.98	1.98	2.11	2.40	5.09
8. Number gallons milked	2.30	2.41	2.31	2.40	2.70	3.29	7.71
9. Dairy product receipts	-	-	-	-	-	-	-
10. Number farm acres	1.57	1.59	1.44	1.33	1.26	1.22	1.22
11. Number corn acres	1.66	1.78	1.72	1.72	1.92	2.30	5.25
12. Number oat acres	2.10	2.35	2.49	2.83	3.61	4.77	13.74
13. Corn yield	.57	.61	.60	.65	.77	.96	2.41
14. Oat yield	1.33	1.42	1.40	1.52	1.82	2.29	6.07
15. Commercial feed expenditures	-	-	-	-	-	-	-
16. Total expenditures, operator	2.47	2.60	2.50	2.61	2.96	3.61	8.63
17. Total receipts, operator	2.45	2.68	2.78	3.25	4.05	5.30	14.01
18. Net cash income, operator	6.57	7.18	7.32	8.30	10.19	13.17	34.24
19. Net income, operator	-	-	-	-	-	-	-
20. Number hogs sold	2.55	2.58	2.33	2.16	2.05	1.97	1.97
21. Number cattle sold	5.71	6.44	6.87	8.21	10.50	13.94	38.00
22. Number hogs bought	10.05	11.49	12.48	15.22	19.82	26.59	73.92
23. Number cattle bought	8.90	9.95	10.50	12.40	15.72	20.81	56.42

Table 29. Summary of Sampling Unit Efficiencies. Number of Items Most Efficiently Estimated by the Six-Grid Sampling Units, 1938 and 1939

Expenditure, mileage rate and questionnaire length		Sampling unit					
		S_4	S_2	S	2-S	4-S	36-S
Expenditure of $1000							
I 2¢ / 15 min.	1938	6	10	-	-	1	1
	1939	6½	8½	1	-	2	2
II 2¢ / 60 min.	1938	13	3	-	-	1	1
	1939	14	2	-	-	2	2
III 2¢ / 120 min.	1938	16	-	-	-	1	1
	1939	16	-	-	-	2	2
IV 5¢ / 15 min.	1938	1	12½	2½	-	1	1
	1939	4	9	3	-	2	2
V 5¢/ 60 min.	1938	6	10	-	-	1	1
	1939	7½	8½	-	-	2	2
VI 5¢ / 120 min.	1938	11½	4½	-	-	1	1
	1939	12	4	-	-	2	2
Expenditure of $2000							
VII 2¢ / 15 min.	1938	7	9	-	-	1	1
	1939	8	8	-	-	2	2
VIII 2¢ / 60 min.	1938	16	-	-	-	1	1
	1939	15	1	-	-	2	2
IX 2¢ / 120 min.	1938	16	-	-	-	1	1
	1939	16	-	-	-	2	2
X 5¢ / 15 min.	1938	5	11	-	-	1	1
	1939	6	8	2	-	2	2
XI 5¢ / 60 min.	1938	12½	3½	-	-	1	1
	1939	12	4	-	-	2	2
XII 5¢ / 120 min.	1938	12½	3½	-	-	1	1
	1939	14	2	-	-	2	2

Efficiency of Incomplete Matching

By design the 1939 sample was half independent of and half matched with the 1938 sample. The problem with which we are here concerned is the estimation of the efficiency with which this incompletely matched sample estimates item means in 1939 as compared with one which is wholly independent.

Let the value of an item (per grid) in 1939 be related to its value in 1938 such that we can express the relationships as

$$y = a + bx \qquad (41)$$

where y and x are values of the item for the same grid in 1939 and 1938, respectively, b the coefficient of regression of y on x and a is a constant.

For the population we can write

$$\overline{Y} = A + B\overline{X} \qquad (42)$$

where $\overline{Y}$ and $\overline{X}$ are the true means per grid in 1939 and 1938 respectively and A and B are the population parameters. After the samples are drawn, we would like to know the best possible estimates of $\overline{Y}$ and $\overline{X}$. Using sample data alone the best estimate of $\overline{X}$ is merely the 1938 sample sum of the item divided by the 900 grids of which it was composed. Let this be $\overline{x}$ and let the 1938 mean of the 450 matched grids be $\overline{x}_m$. Furthermore let $\overline{y}_m$ be the mean of the 450 matched grids in 1939, $\overline{y}_u$ the mean of the 450 unmatched grids in 1930 and $\overline{y}$ the overall mean of all 900 grids of 1939. Ordinarily $\overline{y}$ would be used as the estimate of $\overline{Y}$.

But as an alternative, $\bar{y}_m$ could be adjusted if $\bar{x}_m \neq \bar{x}$ such that the adjusted $\bar{y}_m$,

$$\bar{y}'_m = \bar{y}_m + b\ (\bar{x} - \bar{x}_m) \qquad (43)$$

Now the variance of $\bar{y}'_m$, $\sigma^2_{\bar{y}'_m}$ is given by[19]

$$\sigma^2_{\bar{y}'_m} = \frac{\sigma^2(1-\rho^2)}{n}\left[1 + \frac{N-n}{N}\cdot\frac{1}{(n-3)} + \frac{\sigma^2\rho^2}{N}\right] \qquad (44)$$

where N = number in the 1938 sample
n = number out of N which were matched in 1939
σ^2 = true sampling variance of the item in 1939
ρ = true correlation coefficient of the population

The variance of $\sigma^2_{\bar{y}_m}$ is $\sigma^2_{\frac{y}{n}}$. It can be seen that $\sigma^2_{\bar{y}'_m}$ is less than $\sigma^2_{\bar{y}_m}$ if there is any substantial correlation. Let us assume for the moment we have these correlations and therefore adopt $\bar{y}'_m$ as the best estimate of $\bar{Y}$ from the matched portion of the sample.

We have now two estimates of $\bar{Y}$, $\bar{y}_u$ and $\bar{y}'_m$, which are independent of each other, representing the two portions of the sample and differing in variances. Combining the two for the best overall estimate of $\bar{Y}$ we obtain the weighted estimate (weighted inversely as the variances),

$$\bar{y}_w = \frac{\bar{y}'_m \sigma^2_{\bar{y}_u} + \bar{y}_u \sigma^2_{\bar{y}'_m}}{\sigma^2_{\bar{y}_u} + \sigma^2_{\bar{y}'_m}} \qquad (45)$$

having the variance,

$$\sigma^2_{\bar{y}_w} = \frac{\sigma^2_{\bar{y}_u}\,\sigma^2_{\bar{y}'_m}}{\sigma^2_{\bar{y}_u} + \sigma^2_{\bar{y}'_m}} \qquad (46)$$

---

[19] Developed by W.G. Cochran. It is assumed that N is a small portion of the population.

The variance of the unweighted mean $\bar{y}$ in our sample is

$$\frac{\sigma_y^2}{2n} \text{ or } \frac{\sigma_y^2}{900}$$

We compare the variance of the two estimates $\bar{y}$ and $\bar{y}_w$ to determine relative efficiencies of the incompletely matched sample as compared with an unmatched sample. This relative efficiency is given by the ratio

$$\text{Relative efficiency} = \frac{\sigma^2_{\bar{y}}}{\sigma^2_{\bar{y}_w}} \times 100$$

which in our case where both n and N are large and where $n = \frac{N}{2}$, reduces to

$$\text{Relative efficiency} = \frac{4 - \rho^2}{2(2-\rho^2)} \times 100 \qquad (47)$$

Estimated relative efficiencies on a group of items have been computed to show how much the incomplete matching as followed in the sample survey has increased efficiency over unmatched samples in estimating year means. These estimates appear in Table 30.

It is clear that estimates of the 1939 means were substantially improved by the adoption of the above method of estimation (45). If correlations were perfect (± 1) the gain in relative efficiency would be 50 per cent.

The question may now be raised, what would have been the best fraction to match, assuming that the first year's sample had already been taken and that for a given expenditure the best possible estimates

Table 30. Relative Efficiency of the Half Matched Half Unmatched 1939 Sample Compared with That of a Completely Unmatched Sample When Item Means for 1939 Are Being Estimated

Item	Relative efficiency
	(%)
1. Acres in farms	145
2. Corn acres, harvested	145
3. Oat acres, grain	139
4. Barley acres, grain	131
5. Number of swine	137
6. Number of horses	142
7. Number of cattle	140
8. Number of sheep	143
9. Number of chickens	141
10. Receipts from sales of dairy products	136
11. Gross expenditures, operator	131
12. Gross receipts, operator	138
13. Net cash income, operator	122
14. Number of persons on farms	143

of 1939 means were desired? The problem here is to determine how a given expenditure should be made between n matched and m unmatched sampling units, assuming the unit cost of obtaining each is the same.

The best allocation of sampling resources between n and m will be obtained if

$$\frac{\partial \sigma^2_{\bar{y}_m}}{\partial n} = \frac{\partial \sigma^2_{\bar{y}_u}}{\partial m} \qquad (48)$$

Since

$$\frac{\partial \sigma^2_{\bar{y}_m}}{\partial n} = \frac{-\sigma^2(1-\rho^2)}{n^2} \quad \frac{Nn(n-3)-\sigma^2(1-\rho^2)-\sigma^2(1-\rho^2)(N-n)(2Nn-3N)}{N^2n^2(n-3)^2}$$

and

$$\frac{\partial \sigma^2_{\bar{y}_u}}{m} = \frac{-\sigma^2_{\bar{y}}}{m^2}$$

then,

$$\frac{n^2}{m^2} = (1-\rho^2) - \frac{n}{N(n-3)} + \frac{(1-\rho^2) + (1-\rho^2)(N-n)(2Nn-3N)}{N^2(n-3)^2} \qquad (49)$$

But if N is large, σ^2 fairly large and $n > 3$,

$$\frac{n^2}{m^2} = 1 - \rho^2$$

and

$$m/n = \sqrt{\frac{1}{1-\rho^2}} \qquad (50)$$

When $\rho = 0$, $m/n = 1$, that is when there is no year to year correlation matched and unmatched sampling units are equal in sampling information - it makes no difference whether matching is done or not. When $\rho = \pm 1$, however, $m/n \to \infty$, which would indicate that no matching should be done at all - that only unmatched sampling units should be taken. But (50) is an approximation and appropriate only when N is large and $n > 3$. Actually when $\rho = \pm 1$, n must be two in order that the regression can be determined for the adjustment of $\bar{y}_m$. Any further increase in n would yield no more information, hence all further increase in sample size should be with unmatched sampling units.

For illustration, m/n values, that is the optimum allocation ratios of unmatched to matched, have been computed for a set of items shown in Table 31.

Table 31. Estimated Correlation Coefficients and Optimum Allocation Ratios of Unmatched to Matched Grid Sampling Units for a Selected Set of Sample Survey Items

Item	r	$\frac{m}{n} = \sqrt{\frac{1}{1-r^2}}$
1. Farm acres	.9724	4.29
2. Corn acres	.9709	4.18
3. Oat acres	.9368	2.86
4. Barley acres	.8763	2.08
5. Number of horses	.9539	3.33
6. Number of cattle	.9415	2.97
7. Number of swine	.9229	2.60
8. Number of sheep	.9590	3.53
9. Number of chickens	.9476	3.13
10. Receipts from dairy products	.9185	2.53
11. Total expenditures, farm	.8736	2.05
12. Total receipts, farm	.9269	2.66
13. Net cash income, operator	.7759	1.59
14. Number of persons	.9612	3.62

It can be seen that, for the kind of items investigated, roughly 2 or 3 unmatched sampling units should be taken to every one that is matched. For this particular sampling problem the half-and-half sample is not as efficient as one having a smaller portion matched, regardless of what the correlation coefficient may be.

Let us now consider the problem of determining the allocation of sampling units among the three categories of a sample design involving incomplete matching: (a) N, the sample of the first year (b) n, the matched sample of the subsequent year and (c) m, the unmatched sample of the subsequent year. The problem is to find the relationship which N,n and m must hold with each other such that (a) the variance of the sample means is the same each year and (b) that the total of N, n and m is a minimum for given sampling variances. In other words, what is the best allocation of sampling resources between and within years for this kind of sampling design?

Algebraically we want these three conditions satisfied:

(a) $\sigma^2_{\bar{y}} = \sigma^2_{\bar{y}_w}$ (assuming σ^2 is the same both years)

(b) $n = m\sqrt{1-\rho^2}$

(c) $N + n + m$ minimized for a given $\sigma^2_{\bar{y}}$ or $\sigma^2_{\bar{y}_w}$.

It is assumed here that the population variance, σ^2, remains the same both years, that N is large, that $n > 3$, and sampling units are obtainable at equal and constant unit costs.

Then (a) can be written

$$\frac{\sigma^2}{N} = \frac{\sigma^2_{\bar{y}_u} \; \sigma^2_{\bar{y}_m}}{\sigma^2_{\bar{y}_u} + \sigma^2_{\bar{y}_m}} \qquad (51)$$

and since approximately (if N is large and n moderately so)

$$\sigma^2_{\bar{y}'} = \frac{\sigma^2(1-\rho^2)}{n} + \frac{\sigma^2 \rho^2}{N}$$

and also $\frac{\sigma^2}{\bar{y}_u} = \frac{\sigma^2}{m}$

then $$\frac{\sigma^2}{N} = \frac{\frac{\sigma^2}{m}\left[\frac{\sigma^2(1-\rho^2)}{m} + \frac{\sigma^2\rho^2}{N}\right]}{\frac{\sigma^2}{m} + \frac{\sigma^2(1-\rho^2)}{n} + \frac{\sigma^2\rho^2}{N}} \qquad (52)$$

which when $m\sqrt{1-\rho^2}$ is substituted for n finally reduces to

$$\frac{m}{N} = \frac{-(1-\rho^2+\sqrt{1-\rho^2}) - \sqrt{(1-\rho^2-\sqrt{1-\rho^2})^2 + 4\rho^2\sqrt{1-\rho^2}}}{2\rho^2} \qquad (53)$$

For several values of ρ and for N = 1000, computed optimum values of m and n appear in the following table.

	$\rho = 0$	$\rho = \pm .5$	$\rho = \pm .9$	$\rho = \pm .98$	$\rho = \pm 1.0$
N	1000	1000	1000	1000	1000
m	500	498	443	349	0
n	500	431	193	70	2
Total	2000	1929	1636	1419	1002

Where $\rho = o$, m and n need not be 500 each - it is necessary only that m + n = 1000. Apparently we can conclude that if item year-to-year correlations are rather high (and known beforehand) considerable gain can be obtained by incomplete matching in the manner just considered. By referring to the estimated ρ's given in Table 31, it can be seen that since year-to-year correlations vary quite a lot among items incomplete matching would vary in efficiency according to the item. Farm and corn acres would have worked well whereas "Operator's net cash income" would have done only moderately so (about 10 per cent fewer sampling units being required.)

SUMMARY AND CONCLUSIONS

1. A number of the problems of sampling farm facts by means of two sample surveys taken in Iowa have been studied statistically. These were essentially problems of sampling efficiency and the detection and measurement of biases and other errors in the data.

2. A sample survey of 800 farms provided estimates for the state of Iowa which were in many cases as accurate or even more accurate than corresponding information provided by the Federal Census, Iowa State Farm Census or Federal Agricultural Marketing Service. On the other hand, it was found that some items cannot be accurately estimated from a small (800 farm) sample survey, but these items are in many cases only of minor importance.

3. The sampling methods used in these experimental surveys were found to be not only relatively free of bias but also satisfactorily efficient. Certain modifications, however, have been recommended. Some principles have been suggested for modifying size of sampling unit for maximum efficiency when certain cost situations are given. Another feature of the sampling method is that it provides a basis for making unbiased estimates of total number of farms and total land in farms in any desired area and, therefore, is independent of any other source of information.

4. It has been found that wide geographical distribution of sampling units (that is, geographical stratification into small areas) substantially reduces sampling error. Stratification by tenure group would bring only small gains if any.

5. A method has been proposed for determining the best size of sampling unit for given cost situations and for given expenditure levels. It was concluded that the quarter-section grid is an efficient sampling unit under widely varying circumstances. For investigations requiring very short interviews the half-section grid appears to have important advantages. For general inquiries (such as the census, for example), large blocks such as townships appear to be very inefficient sampling units.

6. Matching samples has proved to be an efficient method of measuring differences between years. Compared with samples taken independently each year, matched samples are from 2.5 to over 20 times as efficient, depending upon the item.

7. Matching as a special case of double sampling has been investigated. As an example of this case: A large sample is taken for a base year; in the subsequent year a small sample is taken at random from the large sample. Precise estimates of the mean of the second year are desired. If there are any correlations for items between the two years, the estimates for the second year can be somewhat improved over that obtainable from small sample considered independently. When the relative variances of these adjusted estimates were compared

with the unadjusted for a selected group of items, it was found that substantial increase in sample information was gained by the matching technique.

8. Sampling errors for a 900 quarter-section sample have been found to range from 2.4 to 14.2 per cent of the grid mean, depending on the item investigated. Number of sheep on farms and number of hogs bought were the most difficult to sample (having highest sampling errors). Sampling errors ranged most generally from 3 to 4 per cent for these samples.

9. Biases and other errors have been detected and measured. The more important biases have been found to be chargeable to the interview method rather than to sampling method. Receipt items were seriously biased (as high as 50 per cent). Failure to remember inventories of a year ago accounted for biases of 10 to 20 per cent. Reports given the Iowa assessors for the annual state census were for some items quite different than those given the survey enumerators. Assessors obtain about a bushel per acre less corn yield than the sample surveys.

Since these biases are not removed by taking larger samples it indicates that improved accuracy is attainable only by improving one's knowledge on the nature and extent of this bias or by eliminating it. We believe that both methods should be tried.

10. Random variations in the reports given the assessor and sample survey enumerators were found and measured. After removing

effects of farm and bias, if any, these variations (considered to be more or less random) measured as standard deviations in percentages of the means ranged from 6 to 88 per cent for those farmers having some of the item. If this is really a measure of the random inaccuracies in interview data then it appears that here is an important source of error. The coefficients of variation of the farm population (stratified) from which the survey sample was drawn ranged from 75 to 425 per cent. It suggests that if attention was given to the problem of minimizing these inaccuracies the precision of sample estimates could be increased somewhat.

11. A high degree of constancy has been found in the distribution of variance in the farm population, although certain shifts were noted. These shifts were such that the desirability of a high degree of geographical stratification was made even more evident.

12. The adoption of the method of repeated visits during the year combined with the use of some simple kind of bookkeeping system would seem to hold promise of minimizing some of the more serious errors in the economic items. Perhaps only two or three visits during the 12 months will be necessary. All pertinent information from previous visits should be available during the subsequent visits for memory aids to both interviewer and interviewee.

ACKNOWLEDGMENTS

The author wishes to acknowledge the assistance, suggestions and criticisms of the following: Gerhard Tintner, George W. Snedecor, T.W. Schultz, C.F. Sarle, W.G. Cochran, C.P. Winsor, Wylie D. Goodsell, C.W. Crickman, Arval Erikson, A.J. King and Norman V. Strand.

LITERATURE CITED

Anderson, Oskar.
1934. [Description of the Bulgarian Agricultural sample census] Bulletin de Statistique, Publ. Direction Gen. de Statistique. de Bulgarie, No. 8. (Summary in French).

Blankenship, Albert B.
1940. Pre-testing a questionnaire for public opinion poll.
A Sociometry 3:263-269.

Blankenship, Albert B.
1940. The choice of words in poll questions. Sociol. and
B Social Res. 25:12-18.

Bowley, A.L.
1906. Address to the Economic Science and Statistics Section of the British Association for the Advancement of Science. Jour. Roy. Statis. Soc. 69:540-558.

Bowley, A.L.
1913. Working class households in Reading. Jour. Roy. Statis. Soc. 76:672-701.

Bowley, A.L.
1926. Measurement of the precision attained in sampling. Bulletin de l'Institut International de Statistique, 22(1):6-62.

Cochran, W.G.
1939. The use of the analysis of variance in enumeration by sampling. Jour. Amer. Statis. Assoc. 34:492-510.

Cochran, W.G.
1942. Sampling theory when the sampling units are of unequal sizes. Jour. Amer. Statis. Assoc. 37:199-212.

Finkner, A.L., Monroe, R.J. and Morgan, J.J.
1943. North Carolina farm employment statistics based on sample data. (To be published).

Gallup, G.H. and Rae, S.F.
1940. The pulse of democracy; the public-opinion poll and how it works. New York, Simon and Schuster.

Gini, Corrado.
1928. Une application de la methode représentative aux matériaux des dernier recensement de la population italienne. Bulletin de l'Institut International de Statistique, 23(2):198-215.

Hendricks, W.A.
1942. The precision of a weighted average. U.S. Dept. of Agric. Agric. Marketing Service. Crop Reporting Procedure Circular No. 6. 8 p. mimeo. July 2, 1942.

Holmes, Irving.
1939. Research in sample farm census methodology. Part I: Comparative statistical efficiency of sampling units smaller than the Minor Civil Division for estimating year-to-year change. U.S. Department of Agriculture, Agricultural Marketing Service.

Hopkins, J.A.
1941. An experiment on the accuracy of farm survey data. Jour. Farm Econ. 23:492-496.

Hoijer, Ernst.
1937. Organization of the 1935-36 extraordinary census of Sweden. Iowa Agric. Exp. Sta. Mimeo. Series, No. 3702.

Jebe, E.H.
1941. The classification and analysis of types of farms in Iowa. Unpublished M.S. Thesis. Ames, Ia., Iowa State College Library.

Jensen, A.
1926. A Report on the representative method in statistics. Bulletin de l'Institut International de Statistique. 22(1):355-376.

Jensen, A.
1926. B The representative method in practice. Bulletin de l'Institut International de Statistique. 22(1):377-436.

Kiaer, A.N.
1895. Observations et experiences concernant les dénombrements répresentatifs. Bulletin de l'Institut International de Statistique. 9(2):176-183.

King, A.J. and Simpson, G.D.
1940. New developments in agricultural sampling. Jour. Farm Econ. 22:341-349.

King, A.J. and McCarty, D.E.
1942. An objective method of sampling wheat fields to estimate production and quality of wheat. U.S. Dept. of Agric. Tech. Bull. 814:1-87.

Mahalanobis, P.C.
1940. A sample survey of the acreage under jute in Bengal. Sankhya. 4:511-530.

Morgan, J.J.
1938. Enumerator bias in census data. Proceedings of Conference of Agricultural Statisticians at St. Louis, Missouri. U.S. Dept. of Agric., Agric. Marketing Service. Mimeo. pp. 36-40.

Neyman, Jerzy.
1934. On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. Jour. Roy. Statis. Soc. 97:558-606. Discussion, pp. 607-625.

Neyman, Jerzy.
1938. A contribution to the theory of sampling human populations. Jour. Amer. Statis. Assoc. 33:101-116.

Roper, Elmo.
1940. Sampling public opinion. Jour. Amer. Statis. Assoc. 35:325-334.

Sarle, C.F.
1932. Adequacy and reliability of crop-yield estimates. U.S. Dept. of Agric. Tech. Bull. 311:1-137.

Schumacher, F.X. and Chapman, R.A.
1942. Sampling methods in forestry and range management. Durham, North Carolina, Duke University. p. 213.

Smith, H. Fairfield.
1938. An empirical law describing heterogeneity in the yields of agricultural crops. Jour. Agric. Sci., 28:1-23.

Social Science Research Council.
1928. Research method and procedure in agricultural economics. Vol. 1, pp. 58-90.

U.S. Bureau of Agricultural Economics.
1936. Proceedings of conference on statistical methods of sampling agricultural data. Ames, Iowa.

Warren, G.F. and Livermore, K.C.
1911. An agricultural survey. N.Y. (Cornell) Ag. Exp. Sta. Bull. 295:376-569.

Watson, D.J.
1937. The estimation of leaf areas. Jour. Agric. Sci. 27:474-483.

Winsor, C.P. and Clarke, G.L.
1940. Statistical study of variation in the catch of plankton nets. Jour. Marine Res., 3:-1-34.

Yates, D. and Zacopanay, I.
1935. The estimation of the efficiency of sampling, with special reference to sampling for yield in cereal experiments. Jour. Agric. Sci. 25:545-577.